TÉCNICO LISBOA

Benchmarking the quality of care in hospitals:

Identifying best practices and opportunities for improvement

Ana Carolina Vilela da Cunha

Thesis to obtain the Master of Science Degree in

Industrial and Management Engineering

Supervisor: Prof. Diogo Filipe da Cunha Ferreira

Examination Committee

Chairperson:	Prof. António Sérgio Constantino Folgado Ribeiro
Supervisor:	Prof. Diogo Filipe da Cunha Ferreira
Member of the committee:	Prof. Maria Isabel Craveiro Pedro

November 2023

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

As I reflect upon and express gratitude at this pivotal moment marking the end of my academic journey, I feel a profound need to convey my deep appreciation to some key individuals who made this achievement possible.

First and foremost, I wish to extend my sincere gratitude to my advisor, Professor Diogo Cunha, for the patience, dedication, and guidance, as well as the collaboration in resolving doubts and problems that arose throughout the completion of this work.

To my friends, who have been consistently present, offering laughter in moments of relief and solace in times of stress, I thank you from the core of my being. Your friendship has been a steadfast pillar and a source of strength.

A special acknowledgment goes to my boyfriend. Your patience, understanding, and unwavering support have been invaluable. Through every challenge and victory, having you by my side made all the difference.

And most notably, I wish to extend my heartfelt gratitude to my parents. Their unwavering support, belief in my potential, and unconditional love have been the cornerstone upon which every step of this journey was built.

This work stands as a testament not just to my effort, but also to the influence and backing of each one of you. I thank you all from the bottom of my heart.

Abstract

Inaugurated in 1979, the National Health System (NHS) of Portugal faces significant challenges in the efficient management of its resources, exacerbated by the rising demand associated with the aging population and the increasing incidence of chronic diseases. Due to the high health expenditures, reaching 11.2% of the GDP in 2021, there is an imperative to reduce costs without compromising the quality and accessibility of services, demanding innovations in management and a continuous evaluation of hospital performance. The literature reveals that most research in the health area uses Data Envelopment Analysis (DEA) and focuses on technical efficiency, with little attention to the use of composite indicators (CIs) in hospital quality benchmarking studies. This research employs the Benefit of Doubt (BoD) methodology, accommodating both desirable and undesirable indicators. Performance analysis focuses on 37 public entities, drawing on recent data about Portuguese public hospitals (2019-2021) to understand the factors contributing to these entities' performance. The results of this analysis indicated variations in the performance of the units, revealing a decrease in the average reference values over the period under analysis. Furthermore, the study demonstrated the significant effect of sociodemographic indicators on the performance of these organizations. Finally, the study suggests mitigation measures and continuous improvement strategies, namely the redistribution of tasks and services to optimize service capacity, the continuous assessment of community health needs to align resources, and the establishment of strategic partnerships with healthcare entities and academic institutions, to strengthen organizational capabilities and respond robustly to healthcare challenges.

Keywords: Benefit of Doubt; Performance; Public Hospitals; Composite Indicators; Quality Indicators.

Resumo

Inaugurado em 1979, o Sistema Nacional de Saúde de Portugal (SNS) enfrenta desafios na gestão eficiente dos seus recursos, agravados pela crescente procura devido ao envelhecimento da população e ao aumento de doenças crónicas. Com as despesas de saúde a atingirem 11,2% do PIB em 2021, torna-se imperativo reduzir custos sem comprometer a qualidade e acessibilidade dos serviços, exigindo inovações na gestão e uma avaliação contínua do desempenho hospitalar. A literatura revela uma predominância de investigações na área da saúde que utiliza Data Envelopment Analysis (DEA) e foca na eficiência técnica, existindo pouca atenção relativamente ao uso de indicadores compósitos em estudos de benchmarking de qualidade hospitalar. Esta pesquisa emprega a metodologia Benefit of Doubt (BoD), integrando indicadores desejáveis e indesejáveis. A análise de desempenho foca-se em 37 entidades públicas, recorrendo a dados recentes sobre os hospitais públicos portugueses (2019-2021), para compreender os fatores que contribuem para o desempenho destas entidades. Os resultados desta análise indicaram variações no desempenho das entidades, revelando uma diminuição nos valores médios de referência ao longo do período analisado. Além disso, foi demonstrado o efeito significativo de indicadores sociodemográficos no desempenho dessas organizações. Por fim, o estudo sugere medidas de mitigação e estratégias de melhoria contínua, nomeadamente a redistribuição de tarefas e serviços para otimizar a capacidade de serviço, a avaliação contínua das necessidades de saúde da comunidade para alinhar recursos, e o estabelecimento de parcerias estratégicas com entidades de saúde e instituições académicas, para fortalecer as capacidades organizacionais e responder aos desafios da saúde.

Palavras-Chave: Benefit of Doubt; Desempenho; Hospitais Públicos; Indicadores compósitos; Indicadores de qualidade.

Table of contents

List of tables

List of figures

List of acronyms

ACES	Agrupamentos de Centros de Saúde (Groupings of Health Centers)
ACSS	Administração Central do Sistema de Saúde (Central Administration of National Health System)
AHP	Analytic Hierarchy Process
ARS	Administração Regional de Saúde (Regional Health Administration)
BoD	Benefit of Doubt
BSC	Balanced Score Card
CH	Centro Hospitalar (Hospital Center)
CHU	Centro Hospitalar Universitário (University Hospital Center)
CI	Composite Indicator
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
DRG	Diagnostic Related Group
ED	Emergency Department
EPE	Entidade Pública Empresarial (Public Business Entity)
GDP	Gross Domestic Product
HOBE	Hospital Benchmarking
INE	Instituto Nacional de Estatística (National Statistics Institute)
IOM	Institute of Medicine
KPI	Key Performance Indicators
LSR	Least-Squares Regression
NHS	National Health System
NPM	New Public Management
OECD	Organization of Economic Cooperation and Development
PCA	Principal Component Analysis
PFA	Production Frontier Approach
PPP	Public-Private Partnerships
QI	Quality Indicator
SA	Sensitivity Analysis
SFA	Stochastic Frontier Analysis
SNS	Sistema Nacional de Saúde (National Health System)
TFP	Total Factor Productivity
TMRG	Tempo Máximo de Resposta Garantido (Guaranteed Maximum Response Time)
UA	Uncertainty Analysis
ULS	Unidade Local de Saúde (Local Health Unit)
US	United States
VIF	Variance Inflation Value
VRS	Variable Returns to Scale

1. Introduction

In this chapter, it is provided an overview of the subjects and ideas covered in this dissertation. It includes an explanation of the problem that will be held, the motivation behind this research and outline the objectives. Furthermore, the research questions and hypotheses that will be raised in this study are explained. Lastly, the structure of the dissertation is outlined.

1.1 Problem contextualization

Health care for Portuguese citizens is a right guaranteed by the National Health System (NHS). Created in 1979, it is a universal public system that tends to be free, considering the economic and social condition of the population, established through Law nº 56/79, published on September 15, 1979. It is fundamentally managed in a decentralized and participatory manner (Article 64, point 4 of the Constitution of the Portuguese Republic, 1976), which means that health services are managed at regional and local levels, involving the participation of the community and health professionals. This allows for a more personalized and responsive healthcare approach, as it considers the specific needs and characteristics of each region and community (Afonso & Fernandes, 2008).

The NHS is under the responsibility of the Portuguese Ministry of Health, who, via potential contracts, provides funding to hospitals, primary care facilities, community health centers, and continuing care facilities. This system is responsible for providing a wide range of health services, including preventive and curative care, hospital care, primary care, mental health care, dental care, and rehabilitation services, and also supports research and innovation in the field of health. Furthermore, it is one of the oldest healthcare systems in the world and operates like a Beveridge system, similar to the system devised in 1948 by William Beveridge, the social reformer who created the UK's National Health Service. In essence, this means that the NHS is a tax-funded healthcare system where the government provides medical services to all residents, regardless of their ability to pay. This configuration demands for the effective allocation of resources to ensure that everyone has access to health services, striving for a balance between costs reduction and improvement of the quality of services provided (Matos et al., 2021; Pereira et al., 2021a).

Since the beginning of the 21st century, when many nations – especially those in the European Union (EU) – were hit by an economic crisis, it has become necessary to re-evaluate the way in which resources are being allocated and managed, in order to ensure that investments in health continue to be aligned with the needs of the population. Therefore, the NHS is no exception, as it faces significant challenges in the efficient management of its resources due to the complexity of its internal networks and alternating political changes in the country (Pereira et al., 2021b). As pointed out by Nunes (2014), the major problem in the health sector in Portugal is not just the apparent underfunding, but in fact the way the resources are used. It is a current reality that health systems still face some resource constraints due to the financial and economic crisis of recent years and the pressure exerted by technological and pharmaceutical innovation (Ferreira et al., 2018b). Because of this, solutions to promote and improve sustainability are increasingly sought in the performance and adequate management of the scarce resources available (Ferreira et al., 2018c). Thus, this topic arises from two sides. Firstly, on the supply side, that is, through technological development and therapeutic innovation, even with uncertainties regarding its cost-benefit ratio. Secondly, on the demand side, with increasingly aging populations and an increase in the incidence of chronic diseases (Jacobs et al, 2006; Simões & Marques, 2011).

In recent years, with a lack of resources and an aging population, there has been an escalation in the demand for healthcare services, what puts pressure on Portuguese hospitals to offer high-quality care, despite being limited by financial constraints. According to Instituto Nacional de Estatística (INE), which stands for "National Statistics Institute", the system is under excessive pressure, with current health expenditure growing by 12.2% in 2021, reaching 11.2% of GDP (Gross Domestic Product), the highest level of the series available starting in 2000, which is among the highest in the European Union. In 2023, the weight of the health sector in the state's primary expenditure is amplified to 13.9%, a situation that is aggravated by the fact that these variations cannot be directly interpreted as proportional to the assistance capacity of the NHS. The government's healthcare system has faced criticism, leading to a decline in public opinion and, as all Portuguese residents access public health care through tax contributions, their perspectives are crucial. In this way, the search for quality has become a central objective in most health systems, since hospitals are under great pressure to perform well in their activities, both in the use of resources and in the quality of the results obtained (Arocena & García-Prado, 2007).

1.2 Motivation

In recent years, Portugal has undergone some health reforms as a result of political changes and, more recently, as a result of the financial crisis, from which Portugal is currently recuperating. Enhancing the effectiveness of the healthcare system was one of the follow-up actions taken, but, for reasons of health, doing so should not come at the expense of access and quality. Instead, if at all possible, efficiency and quality improvements should be made simultaneously to the healthcare system (Nunes & Ferreira, 2018).

Given that healthcare expenditure represents a substantial proportion of total public expenditure, there is also enormous pressure to cut costs. However, this cost reduction in the healthcare sector must be done without sacrificing the volume and quality of services provided (Flokou et al., 2017). Thus, one method to make up for budget cuts and preserve, or even raise, the quality of the healthcare services provided is through identification and elimination of possible sources of inefficiency. For instance, it is clear that once inefficiency is eliminated (or at least reduced), the saved resources might be employed to fund other aspects of the healthcare system, such as higher-quality patient care, innovative technology, staff training, etc. (Flokou et al., 2017). Nevertheless, despite the efforts made to optimize the use of available resources, Portuguese hospitals still face several issues that impact the quality of care, including long waiting times, inadequate staffing levels, and limited access to essential medical technologies [1]. Therefore, for their organizations to perform better, administrators are interested in learning about the features of high-performing hospitals in the sector, in order to identify the common characteristics of these hospitals, which may assist in discovering possible causes of low performance

[1] Expatica (2022). The healthcare system in Portugal. Expatica. https://www.expatica.com/pt/healthcare/general-healthcare/the-healthcare-system-in-portugal-106714/, accessed on October 6, 2023.

(Kooreman, 1994). The only way to practice management and make decisions based on facts is to measure hospital activity, carrying out an evaluation process that allows prioritizing opportunities for improvement and evaluating the performance of what has been done (Castro et al., 2014). Therefore, the main motivation for this investigation stands out: the search for actions and behaviors that can guarantee citizens the highest quality of care in NHS hospitals, a fundamental right that they finance through their tax contributions.

At the NHS level, hospitals are the health care providers with the greatest weight in expenditure, being above 50% (Matos et al., 2021). As the infrastructures that most centralize the provision of healthcare, these are the elements whose performance assessment becomes most prudent (Canilho, 2019). While smaller units, such as individual services or physician practices, can provide a more accurate measurement due to their homogeneous nature, they cannot provide a comprehensive understanding of health system performance. Unlike these smaller units, hospitals incorporate a diverse range of services and operations, covering a broader spectrum of the healthcare sector (Canilho, 2019). Therefore, this study defined the evaluation of hospitals as representatives of the quality of health services, allowing a more holistic view of the quality of the health system.

1.3 Objectives

The fundamental objective of public hospitals, as public service institutions, is to meet the needs of patients without seeking financial profit. Therefore, this research aims to analyze whether these objectives are achieved by public health entities in Portugal, focusing on evaluating the quality of health services.

In this sense, this analysis arises from the creation of a composite indicator (CI), through the aggregation of several individual indicators, based on the analysis of various aspects related to the quality of services provided.

Furthermore, it is intended to understand whether there is a connection between the quality of healthcare and hospital environments in Portugal, so this study will also analyze whether various external factors have an influence on CI values.

Finally, through the results obtained, a comparative analysis between the different entities will be carried out, in order to identify the hospitals with the best performance and discern the factors that contribute to their success, as well as outline viable strategies for continuous improvement.

1.4 Research questions and hypothesis

To fulfill the stated research objectives, the analyzes carried out in this research work aim to obtain answers to the following research questions, accompanied by their respective null (H0) and alternative (H1) hypotheses:

R_1) Which hospitals are best rated in relation to the quality of health services? How did the performance of Portuguese public hospitals evolve from January 2019 to December 2021?

R_2) Can sociodemographic indicators justify the evolution of performance in Portuguese public hospitals?

H0(R_2): Sociodemographic indicators do not influence the evolution of the performance of Portuguese public hospitals.

H1(R_2): Sociodemographic indicators influence the performance evolution of Portuguese public hospitals.

With the first question it is intended to carry out a comprehensive comparison of the quality of health services provided by 37 different health organizations in Portugal. To this end, the aim is to aggregate some performance indicators into a CI that can give a comprehensive aggregated measure of the quality of hospitals. The focus in this issue is on discerning the hospital entities that present superior performance and understanding the reasons that led to such a result.

Concerning the second question, a regression analysis will be conducted, based on a hypothesis suggesting significant impacts of sociodemographic factors on hospital performance. Thus, this result may suggest that other factors, such as hospital management practices, organizational culture, and other internal factors, may have less influence on hospital performance than previously considered. This will be further explained in section 4.5.2. This interpretability analysis of performance indicators against socio-demographic factors is crucial. Firstly, it sheds light on elements affecting hospital performance, aiding policymakers in informed decision-making regarding resource allocation and policy development. For example, identifying factors linked to underperformance allows targeted interventions. Secondly, this analysis helps broaden understanding concerning health service equity discrepancies, pinpointing areas lacking equitable healthcare provision.

1.5 Dissertation structure

This document comprises six chapters organized as follows.

Chapter one contextualizes the issue, shedding light on high NHS health expenditure and the increasing demand for health services, underlining the importance of quality assessment. It also presents the motivation, objectives, and research questions, ending with the document's structure.

Chapter two explores the literature review, defining key concepts, such as "benchmarking" and "quality in healthcare", as per author references. It also differentiates quality and efficiency, as most health studies focus on the latter, and discusses commonly used methodologies in such analyses, comparing them and citing relevant studies.

The third chapter outlines the methodological framework, describing the Data Envelopment Analysis (DEA) and Benefit of Doubt (BoD) methodologies along with their distinct features.

The fourth chapter presents the case study of this dissertation, providing an overview of Portugal's NHS and related reforms. It specifies the main aspects of the BoD model to be utilized, including the sample, analysis period, data collection and processing, and the variables and indicators to be employed.

In the fifth chapter, the entities’ performance results are displayed, interpreted, and discussed per the objectives, also presenting mitigation measures for the continuous improvement of services. Moreover, answers to the initial research questions are established based on the findings.

Lastly, the sixth chapter presents a conclusive synthesis of the study, displaying the main conclusions, encountered limitations, and suggestions for potential future work.

2. Literature review

This chapter presents an exhaustive review of the literature carried out in the area of health quality, in which numerous articles were reviewed, 52 of which were selected for the subsequent selection and definition of the variables to be included in this study.

The initial section of this chapter presents a concise and well-defined theoretical framework that outlines the definition of the most employed concepts in this study, "benchmarking" and “quality in healthcare”, based on what the authors alluded to in the articles examined. Next, due to the prevalence of articles that favor efficiency over quality, a crucial distinction between these two concepts will also be explored, providing a lucid understanding of their respective meaning in the evaluation of health services. In addition, this research will address the methodologies frequently used by authors in the health field, allowing the identification of the most common and effective approaches used to evaluate the quality of health services. Subsequently, a comprehensive comparison will be drawn between the two most widely used measures, the non-parametric technique of DEA and the parametric technique of Stochastic Frontier Analysis (SFA), highlighting their respective advantages, disadvantages, and areas of application. The following section focuses on the methodology that will be implemented in this dissertation, highlighting the justification for this choice. Regarding this method, reference is made to international studies that used this methodology to evaluate the performance of hospitals around the world, and then studies focused on the Portuguese context stand out, providing a solid basis for the application of this methodology in this study.

The purpose of this chapter is to convey the conclusions of the different authors, as well as the limits they consider inherent to the formulation and application of their research models. The comprehensive data extraction table, containing all articles initially reviewed to gather the variables employed in this research, together with the case studies, sample information, variables used, and main conclusions, can be found in table 8, in the Appendix.

2.1 Definition of benchmarking

Benchmarking is a process of identifying and comparing practices developed by organizations, in relation to those that are considered a reference, within and outside the same sector of activity, helping to measure and improve an organization's performance. Generally speaking, it is an activity of comparing organizational processes, which involves finding suitable comparison partners, understanding comparative strengths and weaknesses, identifying areas for improvement, identifying gaps, and setting goals (Collins-Fulea et al., 2005). Therefore, this technique is based on three principles: quality maintenance, customer satisfaction, and continuous improvement, with the main objective of achieving a level of excellence, through identified good practices (Cole, 2009).

Benchmarking plays a crucial role in quality improvement activities and six sigma for finding best practices, recognizing shortcomings, and defining goals (Burstin et al., 1999). Being first used by Xerox in the 1970s, it has grown to be a widely used corporate strategy to assist managerial decision-making and continuous process improvement (McNair et al., 1992). Thus, it is clear to state that benchmarking is the process of “finding and implementing best practices” (Camp et al., 1994). As this search for best practices can be internal or external to the company, there are, amongst others, two important types of benchmarking: internal and competitive. Internal benchmarking is the comparison of identical internal operations, while competitive benchmarking is the comparison of an entire or partial organization to its competitors (Camp et al., 1994). In the field of health care, Mosel and Gift (1994) introduced the idea of collaborative benchmarking, which combines internal benchmarking with external (or competitive) benchmarking, and expands the definition of "internal" to include multiple companies that are part of a collaborative network.

According to Braillon et al. (2008), benchmarking in the healthcare system initially emerged in 1990 with the requirements of the Joint Commission on Accreditation of Healthcare Organizations in the United States (US), which defined benchmarking as a monitoring instrument for the impact of governance, management, and clinical and logistical functions. Subsequently, in 2017, the official benchmarking website for Portuguese hospitals was developed by the Central Administration of the Health System (ACSS, abbreviation for the Portuguese words “Administração Central do Sistema de Saúde”), in a project that aimed to raise standards of patient care and the financial and economic performance of institutions affiliated with the NHS.

In most regions of the world, benchmarking methods are expanding in the health sector as a method for reducing costs, enhancing efficiency, and improving quality. According to a study by Solucient on healthcare, if all hospitals in the USA operated as well as the best hospitals, 57000 extra patients would survive each year, 18% fewer medical complications would occur, average hospital lengths of stay would significantly shorten, and $9.5 billion would be saved (Chenoweth, 2003).

2.2 Quality in health (hospital) care

Measuring the performance of hospitals is not an easy task. In fact, it is necessary to consider all possible dimensions of hospital care and the way they interact with each other, in order to achieve the hospital's main objective: improving the patient's quality of life while maintaining financially and environmentally sustainable operations (Caldas & Varela, 2023).

2.2.1 Definition of quality in healthcare

Health quality assessment is a term that was first introduced by Donabedian (1988). However, according to this author it is a tough term to define and attempts to do so in a single line are frequently ineffective (Donabedian, 2005). Therefore, rather than attempting to come up with a general definition, the majority of authors have concentrated on defining quality in terms of the various dimensions of care.

According to the Institute of Medicine (IOM), quality is defined as "the degree to which health care services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge”. The IOM also outlines five essential qualities that define high-quality medical care: fairness, efficiency, timely care, patient focus, and safety. In turn, the World Health Organization (WHO) defines health care quality as “a process for making strategic choices in health systems” (WHO, 2006).

Donabedian (1980) mentions how disastrous it is to measure the quality of health care without first having a precise definition of what is meant by this concept. This author states that assessing quality implies knowing, as a whole, the relationship between its three dimensions: structures (e.g., personnel, equipment, and finances), processes (e.g., consultation, referral, and prescription), and outcomes (e.g., health status and user assessments). According to this author, there are at least three distinct characteristics of high-quality medical treatment: the technical aspects of quality, the interpersonal aspects of quality, and the amenities of care. The technical component of quality relates to how effectively medical knowledge and science are used to the identification and treatment of the condition, the interpersonal aspect refers to the relationship between the patient and the provider, or the responsiveness and attention to detail of the healthcare provider, and the last one, amenities of care, relates to the comfort of the medical institution.

Allen-Duck et al. (2017) conducted research to clarify the definition of quality in relation to healthcare. Thus, the authors identified four key attributes: effectiveness, safety, culture of excellence, and desired results. Therefore, based on these attributes, they defined health quality as "the evaluation and provision of effective and safe care, reflected in a culture of excellence, resulting in the achievement of optimal or desired health". Ideally, it is measured as the improvement in health as a direct result of the care provided, or a measure of the ultimate health outcomes (Ferrier & Trivitt, 2013). In turn, Maxwell (1992) suggested a set of characteristics that may be the most well-known in the UK, defining quality in terms of effectiveness, efficiency, equity, appropriateness, and accessibility. According to a simplified view on healthcare quality put out by Campbell et al. (2003), it may be assessed by considering two key factors: access and effectiveness. This method focuses on whether patients can comfortably and easily get the healthcare they require, as well as if the treatment they receive is of a high caliber and fits their demands.

It is undeniable that the quality of the healthcare services provided affects how well they function in society. Donabedian (1988; 2005) also introduced a comprehensive overview of quality in healthcare, encompassing several elements that contribute to the provision of high-quality care. Such factors include the availability of infrastructure, equipment, and technology. It also includes the activities of clinical and non-clinical staff during the recovery period, including patient safety and eventual treatment outcomes, as well as patient satisfaction and survival rates. Although the success of a treatment is unquestionably an important aspect of quality, it is vital to understand that access to timely and appropriate healthcare services is also crucial for enhancing patients' satisfaction.

The fact that quality has two dimensions or components—technical and perceptual—increases the complexity in analyzing it. Technical quality refers to the extent to which requirements are being satisfied, while perceptual quality is how the consumer "perceives" the quality (Kleinsorge & Karney, 1992). The definition of quality used in this study, taken from Groocock et al. (1988), includes both of these ideas: "The quality of a service is the degree of conformance of all of the relevant features and characteristics of the service to all of the aspects of a customer's need, limited by the price and delivery he or she will accept". This lengthy and disparate list of different criteria inevitably contributes to the confusion about the concept of quality of care. However, it is important to have a clear conceptual understanding of the meaning of this term, as it will influence the types of health policies and strategies that are implemented to improve it. To ensure a complete assessment and credible benchmarking of healthcare services, all these criteria must be considered.

2.2.2 Quality vs efficiency

In the complex scenario of constant evolution of hospital services, the search for excellence is a fundamental priority. This analysis has been a topic of interest for countless researchers and health professionals around the world and, as we delve deeper into the research, a theme that appears recurrently is the relationship between quality and efficiency in the provision of health care. Therefore, to have a comprehensive assessment of hospital performance, both of these concepts must be evaluated.

The main objective of health services is to increase consumers' quality of life and their health status (Ferreira et al., 2018c). According to the WHO (2006), to be of high quality, healthcare must be provided efficiently, in order to optimize resources, avoid waste, and respond promptly to patients' needs. To achieve this, health policies focus on increasing consumer satisfaction and well-being, while trying to sustainably maintain adequate quality (Chang et al., 2011). This provides the basis for creating a more equitable and sustainable health system, with the increasingly sought-after convergence of variables that are not always reconcilable: quality in service provision and efficiency in resource allocation (Nunes, 2014).

According to Hollingsworth (2008), a company is technically efficient when it generates the greatest amount of output from a given quantity of input or, alternatively, when it generates a given output with the least amount of input. When the input mix reduces costs given input prices or, alternately, when the output mix optimizes profits given output prices, a company is allocatively efficient. When a company is overall efficient, it operates on its cost or revenue frontier (Hollingsworth, 2008). Overall, efficiency is made up of both technical and allocative efficiencies. Similarly, Farrell (1957), one of the biggest economists in the twentieth-century, defined efficiency as “the firm’s success to produce the maximum feasible amount of output from a given amount of input, or producing a given amount of output using the minimum level of inputs, where both the inputs and the outputs are correctly measured”.

Efficiency and quality are often thought of as, if not mutually exclusive, at least of there being a trade-off between them (Ozcan, 2009). Previous studies reviewed showed conflicting findings regarding the relationship between these two concepts. Singaroyan et al. (2006) concluded that increasing the quality of medical treatment might not necessarily result in more efficient procedures, while Helling et al. (2006) discovered that quality will increase as efficiency does. In turn, Ferrando et al. (2005) mentions that, with the right principles, hospitals may improve efficiency without compromising the quality of care.

Laine et al. (2005b) examined the relationship between clinical quality and productive efficiency in institutional long-term care for elderly patients and discovered no consistent relationship between both of them. However, it was discovered that long-term care facilities which had high levels of technical efficiency also attained high scores in some quality dimensions (depression medication and treatment). After quantifying technical efficiency and investigating the relationship between quality and technical efficiency, this author discovered a substantial correlation between technical efficiency and what they called "unwanted dimensions of quality".

Despite the importance of incorporating quality as an essential component in efficiency management, many studies often neglect this aspect when measuring efficiency, and several studies that fail to do so recognize this gap as a limitation in their analysis (Arocena & García-Prado, 2007).

This omission can be attributed, in part, to the complexity of defining and quantifying quality, as well as the ongoing discussions around the balance between quality and efficiency, in addition to the lack of consensus on a direct relationship between these two concepts (Bull, 1994; Chang et al., 2011). However, the exclusion of the concept of quality in the cost/efficiency analysis may result in the omission of a critical variable, which could distort any conclusions or inferences based on the data collected. The main concern is that a lack of consideration of quality may lead to a situation in which hospitals that provide lower-quality care appear to be more efficient, while hospitals that provide high-quality care may appear to be using resources less effectively. Therefore, it is extremely important that public health policies cover both hospital efficiency and quality of care in their assessments (Ferrier & Trivitt, 2013).

2.3 Techniques for measuring the quality of healthcare

The growing interest in studying the performance of various entities in diverse domains has led to the emergence of a category of models called frontier estimation techniques. This notion of "frontier" comes into play when evaluating a group of units, allowing the identification of those that present the most favorable results. In essence, these standout values serve as a benchmark or frontier against which all other units must be evaluated and compared.

There are several approaches to carrying out a comparative analysis of performance in health services, and the mathematical models used vary considerably. Within the scope of statistical analysis, there are two types of approaches, each with its own characteristics and application methods - parametric and non-parametric techniques (Arocena & García-Prado, 2007). Parametric methodologies assume a functional form for the efficiency frontier, while non-parametric methodologies estimate the frontier limit empirically (Ferreira et al., 2018a). According to Hollingsworth (2008), DEA has been the most frequently used model for studies of healthcare organizations, followed by SFA methodology. The reasons that support the preferential use of these techniques over others are the fact that they are considered more solid and more suitable for the production process of healthcare units (Gomes, 2015). In addition to these, there are also other types of techniques used to measure performance in the healthcare, but with less expression, such as the Difference in Differences, the translog function, the cost performance index, the deterministic frontier approach, the Malmquist productivity index, and the free disposal hull (Worthington, 2004).

Regarding parametric methods, the most commonly implemented by the academic community in the context of efficiency analysis is SFA. This model was developed with the concern that in the DEA all deviations from the Best Practices Frontier (BPF) would be considered inefficient. This type of approach aims to determine the efficiency of an organization in relation to a theoretical limit, establishing a smooth parametric frontier between efficient and inefficient production. To do so, it assumes a certain data distribution and employs econometric methods to estimate parameters of a production function. However, the defined theoretical frontier presents deviations, the identified causes of which are randomness and technical inefficiency. Therefore, to incorporate these two sources of deviation, the models use two stochastic variables: statistical noise and inefficiency, so the advantage of this technique lies in its ability to distinguish the dimensions of these two effects in the final efficiency measure. Despite the effort to separate the effects of noise and inefficiency, SFA distance functions are parametric and deterministic and, as such, can confound "the effect of omitted variables and measurement errors, as well as possible functional form misspecification" (Jacobs et al., 2006). Moreover, assumptions are also made regarding the distribution of errors, assuming they are independent of each other. In this way, inefficiency is considered to follow a semi-normal distribution, as it only acquires positive values, while, in relation to statistical noise, it is assumed that it follows a normal distribution (with zero mean and constant variance). The SFA technique is based on a uni-input or uni-output model, which can have several explanatory variables and, for this reason, requires a measure of aggregated inputs or outputs to represent the other input or output variables, respectively. Regarding the aggregation of results, this process is carried out in two phases, first estimating the frontier, and then calculating the distance of each unit to that frontier (Canilho, 2019).

In turn, the non-parametric approaches do not estimate particular parameters or make assumptions about the distribution of the data. They are more adaptable and may be used in a broad range of circumstances, even when parametric models' underlying presumptions are compromised. These methods include DEA, Wilcoxon-Mann-Whitney test, Kruskal-Wallis test, and Spearman correlation. Within these, DEA is a technique widely used in the context of efficiency analysis to assess the relative efficiency of decision-making units (DMUs), such as hospitals or other companies. This approach creates a piece-wise linear-segmented efficiency frontier while making few assumptions about the underlying technology, making it less vulnerable to specification error and yet completely immune to random error (Cook & Seiford, 2009).

According to Sherman (1984), the most commonly used methods for assessing efficiency are the following: ratio analysis, balanced score card (BSC), total factor productivity (TFP), regression analysis, production frontier approach (PFA), and DEA. Although ratio analysis can evaluate only one input and one output component, the BSC, TFP, and PFA methods are appropriate for questions with many inputs and one output. In addition to these, there is also Least-Squares Regression (LSR), also known as OLS (Ordinary Least Squares), which is one of the most popular parametric methods in efficiency evaluation, allowing multiple inputs and outputs (Ozcan, 2008). When a technological adjustment to time series data is necessary and when examining economies of scale, the advantages of LSR are realized. However, due to its parametric formulation, it requires a predetermined production function, and it does not specifically identify top performers or inefficient units (Ozcan, 2008). In turn, since it can simultaneously examine multiple elements of investment and production, DEA is best suited for use in the complex environment of medical institutions and the healthcare sector. Due to the fact that this methodology is widely used in the public health system, being a flexible approach, unrestricted in terms of technology, and simple to be used in multi-product scenarios (Ferrari, 2006), it will be used in this work.

2.3.1 Comparison between DEA and SFA methodologies

According to the most recent literature review by Hollingsworth (2008), frontier techniques—primarily DEA but also with some weight SFA —appear to be the dominant methodologies in the evaluation of hospital performance, as shown in Figure 1.

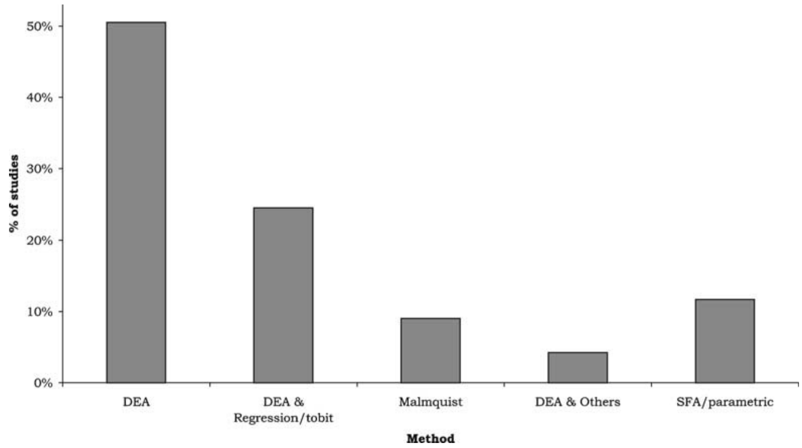

Figure 1- Distribution of methodologies used in studies on health performance (Source: Hollingsworth, 2008).

DEA and SFA stand as two widely employed techniques for evaluating organizational performance. By comparing these two models, applying them to the same variables and the dataset, discrepancies in the efficiency estimates obtained become evident. These differences derive, fundamentally, from the way each technique establishes and configures the efficiency frontier, as well as its methodology for determining the distance between each observation and this frontier (Jacobs et al., 2006). Consequently, despite both methods being centered around the estimation of an efficiency frontier constructed to compare units, each method exhibits distinct advantages and limitations.

DEA operates as a non-parametric, locally segmented index that constructs a linear efficiency frontier. Notably, it requires fewer assumptions about the underlying production technology structure when contrasted with SFA, thereby endowing it with greater robustness. In the context of DEA, the efficiency frontier is solely determined by the available dataset, devoid of the need for theoretical considerations. Additionally, this model offers the flexibility to incorporate various assumptions concerning frontier relationships, returns to scale, input and output availability, as well as the convexity of the production function when information is accessible (Amado & Santos, 2009). This inherent characteristic makes DEA less vulnerable to specification errors, as it does not risk conflating inefficiency with an erroneous specification of the model's functional form — a scenario that can occur when utilizing the SFA method (Castro, 2011). Thus, in instances where the primary objective of performance analysis lies beyond profit maximization, DEA is more frequently applied since it does not impose such constraints. Nevertheless, it is more susceptible to inaccuracies in data and the presence of outliers, although there are ongoing studies aimed at mitigating this issue (Vasquez, 2012). Conversely, SFA adopts a parametric approach that relies on assumptions about both the functional form of the efficiency frontier and the error distribution. It estimates variables that elucidate the relationship between inputs and outputs through a statistical model, implying a precise definition of the functional form and an assumption of an appropriate error distribution. Consequently, if these underlying assumptions are not met, SFA efficiency estimates may exhibit bias or inconsistency.

The DEA technique offers the advantage of accommodating multiple inputs and outputs (Carrilo & Jorge, 2017), a flexibility that the SFA lacks. In practical terms, DEA compares the input-output performance of a hospital against a limited set of efficient peer units to estimate efficiency indices, a method that is particularly sensitive to variations in the data, making it responsive to changes. In contrast, the SFA methodology employs estimated average parameters to derive efficiency estimates, enhancing the robustness of the efficiency estimate against the potential influence fluctuations or outliers in hospital data at an individual level (Castro, 2011).

Nonetheless, one limitation associated with DEA is its inability to distinguish between statistical noise and true inefficiency. Since it is a deterministic technique, it does not consider the stochastic variation of the data and assumes that any deviation from the BPF is an inefficiency. In other words, it may attribute inefficiency to random fluctuations or measurement errors in the data, leading to potentially inaccurate efficiency estimations. Consequently, one common criticism of this approach is its sensitivity to random noise (Castro, 2011). Conversely, the SFA method addresses this limitation by incorporating statistical noise into the analysis due to its parametric nature. This difference in methodology implies that SFA is more likely to assign higher efficiency ratings to hospitals than DEA. Additionally, DEA may estimate more challenging efficiency targets compared to SFA, especially when used for target setting, which might happen because DEA calculates the difference between a hospital's actual and desired input-output levels to determine efficiency ratings (Jacobs, 2001).

In the comparison between these two techniques, Jacobs (2001) concluded that each method has its unique advantages and disadvantages, capturing different aspects of efficiency. Furthermore, Linna (1997) discovered that both approaches produced results for personal efficiency scores that were comparable. However, a method that is reliable in both respects is necessary for a credible efficiency study with a minimum informative nature, which is no longer a concern since the gap between these methodologies has been able to reduce. This development may be seen, for instance, in the incorporation of stochastic components into DEA's original framework or in the SFA's increasing capacity to assess multiple outputs at once (Miranda, 2009). O'Neill et al. (2008) claims that DEA makes it easier to study each unit individually in contrast to the SFA, which makes it simpler to examine the sample as a whole. Thus, SFA proves to be a more effective tool for policy analysis, while DEA provides valuable managerial insights for hospital management, aligning with the focus of this research. Therefore, as highlighted by Amado and Santos (2009), when analyzing entities within the healthcare sector, particularly hospitals, DEA emerges as the preferred model. Its appeal lies in its versatility: it can be easily constructed for multiple inputs and outputs, incorporates all available data to define the efficient production frontier for comparison with each production point, and offers flexibility in projecting the BPF to suit the specific problem context. This adaptability allows the BPF to represent either production maximization, input minimization, or a combination of both as needed.

2.3.2 Hospital performance studies that employ the DEA technique

DEA is an approach that is frequently employed in the medical field to evaluate the performance of healthcare organizations, with numerous studies underlining the importance of carrying out this investigation. While examining the healthcare systems of various European countries, there is a prevailing belief that waste levels range from 20% to 30% of healthcare expenditures.[2] Specifically, concerning losses incurred due to fraud and clinical errors, the Organization of Economic Cooperation

[2] Silva, P. (2022). "Desperdício no SNS". Ernst & Young: https://www.ey.com/pt_pt/health/desperdicio-no-sns, accessed on March 13, 2023.

and Development (OECD) estimates that they contribute to more than 6% of healthcare expenditures. Consequently, employing DEA to evaluate hospital performance can have a positive impact on the operations of healthcare institutions, as it enables the identification of areas where these organizations may fall short, and which stand out more positively. Given the imperative need to reduce waste, which, in turn, allows for the more efficient allocation of scarce resources available for critical investments, this methodology proves to be instrumental in optimizing resource allocation and implementing improvement strategies.

2.3.2.1 Performance of hospitals in the world

Since the early 1980s, there has been a growing trend of applying efficiency and performance analysis to healthcare institutions around the world, evaluating their productivity and operational performance. Consequently, in recent years, there has been a substantial increase in the number of studies that use the DEA method for this purpose, as illustrated in Figure 2.

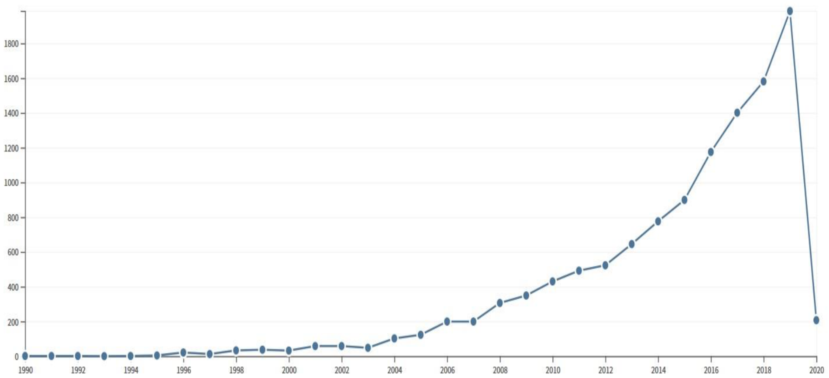

Figure 2- Publications per year that employ the DEA method (Source: Silva et al., 2020).

As per the research conducted by Silva et al. (2020), a majority of studies in the domain of hospital efficiency predominantly employ the DEA, as was also observed in the articles reviewed for this study. Hence, it is crucial to grasp the scope of analysis undertaken by these authors, encompassing both domestic research in Portugal and international studies. Additionally, several studies employ extensions of the DEA methodology, including the BoD approach, which will also be mentioned in the context of this literature review.

The pioneer in the application of DEA methodology within the healthcare sector was Thomas Nunamaker, in 1983. In his groundbreaking study, he assessed the technical efficiency of nursing services in 16 hospital units in Wisconsin, US, spanning the period from 1978 to 1979. Subsequently, Sherman (1984) introduced a more comprehensive approach to measuring hospital efficiency on a global scale. Following this development, the frontier methodology emerged as the most prominent and effective means of evaluating efficiency in the healthcare field (Hollingsworth, 2003). Since then, publications utilizing DEA have proliferated rapidly and continue to be widely employed today.

In 1988, Borden conducted an assessment to investigate how a new reimbursement system, known as the Diagnostic Related Group (DRG), impacted the performance of hospitals in New Jersey, USA.

The findings were subsequently compared with other conventional efficiency measurement methods, including regression analysis and indicator analysis. The study concluded that the variations in hospital efficiency before and after the implementation of the DRG system were not influenced by the new financing mechanism.

Earlier research in the field of DEA often focused on specific characteristics or categories of hospitals, such as teaching and non-teaching hospitals. For instance, in 1996, Magnussen conducted a study to assess the utility of efficiency measures as tools for monitoring and resource allocation, through an analysis of their invariance to changes in the operationalization of hospital production. His analysis included 46 Norwegian acute care, non-teaching hospitals. The findings indicated that changes in the specification of hospital output did not significantly impact the distribution of efficiency scores. However, the choice of output specification did influence how hospitals were categorized and the scaling properties of the technology.

In 2002, Hofmarcher et al. conducted a comprehensive study focusing on the performance of the Austrian healthcare system before the introduction of the DRG system. Covering the years 1994 to 1996, it aimed to evaluate efficiency rankings while employing non-parametric frontier models, utilizing panel data to examine dynamic changes in productivity and efficiency within a specific region of the healthcare sector from Austria. To do this, they used two distinct models to assess the efficiency levels of 31 individual hospital wards, and both models led to the conclusion that the overall changes in efficiency during this period were quite small.

Another study by Grosskopf et al. in 2004 utilized the DEA approach to measure relative technical and scale efficiencies among 254 US teaching hospitals. The aim was to investigate whether market variables and institutional characteristics contributed to performance variations among these hospitals. Through bivariate statistical analysis, the authors found that increased competition was positively correlated with efficiency, instructional commitment, and intensity. Additionally, the study found that commitment to teaching, non-public ownership, and certain affiliation effects had favorable impacts on productivity.

Laine et al. (2005b) conducted a study where technical efficiency was assessed using the DEA technique, and the relationship between quality and efficiency was explored through the application of the Mann-Whitney test and correlation coefficients. The findings from this research indicated a potential link between technical efficiency and certain undesirable aspects of quality within healthcare settings.

In 2006, Kontodimopoulos et al. delved into the efficiency of a group of 17 small-scaled Greek hospitals referred to as hospital centers. The study revealed that Greek hospitals and healthcare facilities in rural areas exhibited notable operational inefficiencies, highlighting the need to address these inefficiencies while ensuring equitable access to healthcare services. Also in that year, Linna et al. (2006) conducted a study using DEA to assess cost efficiency, specifically comparing Finnish and Norwegian hospitals. Their analysis utilized cross-sectional data from 47 Finnish and 51 Norwegian public hospitals in 1999, employing identical definitions for cost and production measures. As a result, the study uncovered significant variations in cost efficiency between Norwegian and Finnish hospitals, partly attributed to differences in input costs and lengths of stay.

In 2007, Arocena and García-Prado analyzed the performance evolution of 20 Costa Rican hospitals during the period from 1997 to 2001. Their empirical investigation aimed to assess whether the implementation of management contracts had an impact on the performance of public hospitals in Costa Rica. The study concluded that the introduction of these contracts had a positive influence on hospital performance, particularly in terms of service quality. Interestingly, smaller hospitals outperformed larger ones, possibly owing to the unique challenges faced by larger institutions in implementing management contracts.

In 2008, Nayar and Ozcan conducted an analysis employing DEA models to assess the performance of 53 non-federal acute care hospitals located in Virginia. Their study focused on evaluating both the technical efficiency and the quality aspects of these hospitals, aiming to investigate whether there was a trade-off between efficiency and quality. To achieve this, they utilized constant returns to scale (CRS) input-oriented models and examined efficiency while considering both quality factors and excluding them from the analysis. Thus, they concluded that enhancing technical efficiency did not necessarily come at the expense of compromising quality in hospital care. In the same year, Lee et al. (2008) studied the relationship between the case-mix specialization index and the efficiency of inpatient hospital services, using a sample of 106 acute care hospitals situated in Seoul, for the year 2004. The findings from their investigation revealed that internal provider qualities had a more significant impact on the specialization of hospitals than external market factors, and hospitals that exhibited higher levels of specialization were also found to be more effective in delivering patient care.

In 2009, a study conducted by Mark et al. aimed to assess the technical efficiency of various nursing units, including general medical, surgical, and combined medical-surgical units. Their analysis revealed that nearly 60% of the nursing units in the sample were operating below optimal efficiency levels. Interestingly, the study suggested that conventional targets for improving efficiency did not yield the most significant benefits. Instead, the study pointed to potential gains in efficiency through the reduction of medication errors and patient falls, highlighting the importance of patient safety measures. Also in that year, Shwartz et al. (2009) conducted a study focusing on VA nursing homes in 1998, where they delved into the assessment of facility rankings using two BoD approaches: linear programming models and DEA. In this research, composite scores were computed for 32 VA nursing homes utilizing both standard and BoD approaches. Additionally, the study investigated the tradeoff between weight adjustment and rankings. The findings of this study provided valuable insights into the assessment of facility rankings and highlighted the potential advantages of employing BoD methodologies in this context. Thus, discussion of the study emphasized the benefits of BoD methods in acknowledging local performance variations, providing facilities with the flexibility to adjust weights, and fostering productive dialogue. On the other hand, DEA approach was noted for its transparency, as it combines quality indicators without normalizing individual scores and identifies benchmark facilities. The study outlined that the choice between these approaches depends on various factors, including the specific characteristics of the facilities, the nature of the indicators used, and the desired level of transparency in the assessment process.

Blank and Valdmanis (2010), the latter being one of the main contributors to DEA-based studies in healthcare focusing on hospitals (Kohl et al., 2019), investigated the efficiency of Dutch hospitals using the DEA method. Their research focused on elucidating the factors that contribute to cost inefficiencies in these hospitals, considering the operating environments of each facility. Thus, they concluded that the intensity of physician involvement was a significant factor in cost inefficiency, presumably as a result of factors, such as physicians substituting hospital resources for their own time, prioritizing patients' well-being, or optimizing their own remuneration. Also in 2010, Garcia-Lacalle and Martin undertook a comprehensive analysis aimed at investigating whether market-driven reforms, designed to enhance efficiency and quality, exerted varying impacts on urban and rural hospitals within the context of a European healthcare system. Their study focused on the Andalusian Health Service in Spain and involved a meticulous examination and comparison of hospital performance, with a specific emphasis on efficiency and perceived quality. The study considered the unique characteristics of rural and urban hospitals, acknowledging their differences in location, size, and patient demographics. Despite these variations, they concluded that the majority of hospitals, regardless of their urban or rural classification, demonstrated a comparable level of efficiency. Interestingly, the study highlighted that large rural hospitals stood out with higher patient ratings in the perceived quality component, and also exhibited remarkable efficiency, particularly in emergency care.

In the subsequent year, Navarro-Espigares and Torres (2011) conducted a study to assess the relationship between efficiency and quality indicators. Their research delved into the dynamics of 27 Andalusian Hospitals during the years spanning from 1997 to 2004, showing that there is no compromise between efficiency and quality objectives, and emphasizing the importance of considering outcome variables that encompass both the activities performed and the level of services delivered in an assessment of the health system's performance.

In 2013, Ferrier and Trivitt conducted a comprehensive analysis using 'double DEA' to investigate the impact of accounting for quality on DEA technical efficiency scores, across a sample of 1,074 US hospitals. They observed substantial inefficiencies within many hospitals and found that adjusting for quality clearly changed the efficiency scores, noticing that the magnitude of this impact varied depending on the specific quality-adjusted method employed. To measure hospital performance in terms of both technical efficiency and quality, and explore the utility of DEA as a benchmarking tool for assessing quality, Nayar et al. (2013) analyzed 371 urban acute care hospitals in the US. According to this study, more hospitals were found to be functioning at the efficiency frontier when both the efficiency and quality aspects of performance were considered.

In 2016, Chowdhury and Zelenyuk conducted an extensive analysis focused on the production performance of 113 acute-care hospitals in Ontario, spanning the years 2003 and 2006. Their study encompassed a diverse range of hospitals across various geographic locations, sizes, teaching statuses, and other key variables. To evaluate efficiency, they employed a non-parametric DEA technique, complemented by bootstrapping and truncated regression methods. The research outcomes revealed that organizational characteristics exhibited a positive correlation with efficiency, suggesting that hospitals with these attributes tended to operate more efficiently. Additionally, the study indicated that smaller, rural, and non-teaching hospitals demonstrated a higher degree of efficiency in their operations.

In 2017, Khushalani and Ozcan conducted a study involving 1,259 U.S. hospitals, to assess whether there were trade-offs or synergies between the efficiency of producing quality healthcare and the efficiency of delivering medical/surgical care. Their analysis aimed to determine if both aspects could be simultaneously improved within the healthcare sector. The findings of their study indicated that improvements in efficiency within both sub-units, quality, and medical/surgical care, exhibited a positive correlation, suggesting that there are no inherent trade-offs between enhancing quality and efficiency in healthcare delivery. Also in 2017, Flokou et al. applied the DEA method to evaluate the efficiency of 107 Greek NHS hospitals over the period from 2009 to 2013. Their study constructed an inter-temporal efficiency map that depicted the technical performance of hospital activities within the context of the Greek NHS. The research suggests that addressing inefficiencies can lead to significant resource savings, which can be utilized for improving patient care, adopting innovative technologies, and investing in staff training.

In more recent years, Kohl et al. (2019) conducted a comprehensive review of over 260 articles that utilized DEA in the context of hospital management. Their analysis categorized the application of DEA into four primary groups, depending on the specific objectives: estimating efficiency, addressing specific managerial inquiries, evaluating a particular health policy, or innovating and implementing new benchmarking methodologies. The results showed that DEA model is the preferred choice for estimating efficiency scores in 80% of the publications, yet its reliability remains unverified. Thus, this study proposes two strategies for advancing the application of DEA in healthcare: the development of methods to establish the reliability of DEA models and the promotion of enhanced communication between researchers and hospital managers.

The most recent study reviewed in this literature analysis focusing on international hospitals, dates from 2022. Conducted by Lin et al. (2022), the study assessed the efficiency of 19 medical centers and hospitals in Taiwan using DEA and TOBIT regression. Notably, the modified EBITDA indicator and self-payment revenues were used to differentiate between foundation hospitals and those with varying operational efficiency levels. The authors concluded that private medical centers outperformed public ones in technical efficiency, pure technical efficiency, and scale efficiency.

2.3.2.2 Performance of hospitals in Portugal

To conduct this research, it was essential to review numerous articles that specifically addressed health organizations in Portugal. In 1999, Dismuke and Sena assessed how the DRG-based new payment structure affected the technical efficiency and productivity of diagnostic technology in Portuguese public hospitals, employing two frontier models: DEA and a parametric one. The parametric model utilized input orientation and an RCE technology, incorporating data from the years 1992 to 1994. Their findings suggested that the newly implemented method had a positive impact on technical efficiency and the productivity of diagnostic technology.

In 2007, Barros et al. analyzed the efficiency and productivity growth of a representative sample of 51 Portuguese hospitals in the period from 1997 to 2004, for which they employed an innovative approach through the use of the directional distance function and the Luenberger productivity indicator. According to the results, the increase in productivity in these hospitals did not occur in this time period, it was only observed in facilities without changes in technical efficiency and advances in technology. Thus, the article suggests that investments in capital accumulation have the potential to increase both technical efficiency and technological change.

The following year, Afonso and Fernandes (2008) carried out a study to estimate measures of hospital technical efficiency. Their research focused on examining the heterogeneity and efficiency resulting from the recent "privatization" of public hospitals. For that, they utilized a dataset containing information from 68 Portuguese public hospitals within the NHS, spanning the years from 2000 to 2005, to analyze productivity growth. The findings of the study revealed that between 2000 and 2004, with a slight decline in 2005, the NHS experienced positive but modest productivity growth levels. Additionally, the study discovered that individual efficiency scores varied significantly from one year to the next and that only seven hospitals maintained their position on the production potential frontier throughout the entire investigation period.

In the same year, Moreira (2008) conducted an evaluation of the impacts of the 2022 reform of the public hospital sector on technical efficiency. This analysis involved comparing the relative performance of company-hospitals with hospitals that remained under public administration, both before and after the entrepreneurialization process. The study considered a small set of 25 company hospitals and 23 control units, and the efficiency measures were developed using the DEA approach. The findings of the study indicated that hospital firms demonstrated higher efficiency levels than hospitals administered by the government.

Moving forward to 2013, Barros et al. authored an article with the objective of examining the statistical characteristics that define hospitals. Their research utilized a sample of 51 hospitals in Portugal, which the authors categorized into three distinct groups, each characterized by specific attributes, performance levels, and financial situations (debts). The findings of the study revealed that a one-size-fits-all government policy would not be suitable for all hospitals, given the variability among the three identified segments. Consequently, the study concluded that latent frontier models were valuable in describing Portuguese hospitals, especially when considering heterogeneity and segmentation. Additionally, the research observed that hospitals serving multiple counties demonstrated efficient operation, and private hospitals outperformed public hospitals in efficiency.

The following year, Castro et al. (2014) introduced an "ideal" model designed to assess the efficiency of the primary Internal Medicine services offered by most hospitals within the Portuguese NHS, focusing exclusively on the production of the four main production lines. The study employed the DEA methodology to evaluate 48 hospitals across Portugal, with the aim of establishing a benchmarking platform for Portuguese hospitals, to assess and improve hospital performance systematically.

In 2016, Portela et al. conducted research to introduce the Hospital Benchmarking (HOBE) platform, which facilitates the comparative analysis of hospital services using individual or aggregated indicators generated through DEA. The authors argue that comparing services within hospitals is a more effective approach than comparing entire institutions. The HOBE tool was developed in collaboration with hospital managers and the ACSS, who provided the necessary data. A total of 80 indicators were defined for 19 clinical services as part of this initiative.

Ferreira et al. (2018a) investigated the scale's efficiency, the optimal scale for hospital clinical staff, and any potential external variables associated with these factors. The study encompassed an annual analysis of 108 hospitals and emphasized the need to optimize the allocation of clinical staff in Portuguese hospitals, aiming to enhance scale efficiency without compromising essential social values. The authors concluded that the hospitals located on the Portuguese coast are operating above their optimal scale, which clarifies a poor distribution of the workforce in the health area. Thus, to address this issue and improve scale efficiency, healthcare policies recommend adjusting the workforce in hospitals that are operating above or below their ideal size, respectively.

In 2021, Pereira et al. (2021b) assessed the efficiency of secondary healthcare providers within Portugal's NHS. They employed a value modelled slacks-based network DEA with a simulation performed on a matrix-type structure. The study incorporated two distinct perspectives, labeled as Perspective 1 and Perspective 2, to encompass both optimal and real-world viewpoints. The analysis was conducted on 27 healthcare organizations that operated as corporate-public companies, denoted as EPE (short for "Entidade Pública Empresarial", in Portuguese). The study concluded that the majority of efficient hospitals were situated in coastal areas, with the most frequently efficient services being inpatient care and operating rooms. Also, that year, Matos et al. (2021) conducted a comprehensive study utilizing the BoD methodology and data spanning from 2013 to 2017 to delve into the factors contributing to poor performance and high indebtedness levels in Portuguese public hospitals. The analysis involved an assessment of the overall performance of these hospitals using a CI that incorporated four key dimensions: access, efficiency and productivity, financial performance, and quality. This analysis indicated that they need to enhance their entire performance, especially in the financial area. Additionally, they observed the compensatory nature of the classic BoD approach but noted that weight restrictions reduced this effect.

Ferreira and Marques (2021) conducted research that aimed to demonstrate that public-private partnerships (PPP) hospitals can provide health services with social performance levels at least as good as public hospitals. They utilized robust benchmarking methodologies and recent data on Portuguese hospitals, covering the period between 2012 and 2017. Employing the BoD methodology, the authors concluded that there is no reason to believe that public hospitals outperform PPPs in terms of social performance, as both groups exhibit high levels of social inefficiency, particularly in areas like punctuality and adequacy of care. Moreover, Pereira et al. (2021a) wrote an article based on this methodology to construct a CI that assesses the performance of 29 Portuguese public hospitals under two perspectives of hospital activity: users and providers. It revealed that there are not many functional public hospitals in Portugal's NHS, which is a major issue that has to be addressed.

Later in 2022, Nunes et al. developed a study for the construction of a useful CI to assess the quality of life of citizens. For this purpose, they analyzed 277 municipalities located in mainland Portugal and applied the BoD methodology. The findings found regional asymmetries with stronger evidence in terms of health and living situations, but homogeneity regarding environmental concerns, education, and culture. The heterogeneity in health and overall quality of life seems to be partially explained by geographic factors.

Furthermore, the same authors conducted a study in that same year, employing a DEA model of network data related to efficiency and effectiveness, in order to evaluate the performance of hospital units before and during the COVID-19 pandemic. In the early stages of the pandemic, they noticed regular declines in efficiency, which were then followed by recovery to levels above pre-pandemic values. They concluded that, despite the public health care professionals' outstanding resilience and the Portuguese State's lack of readiness for a pandemic like COVID-19, the efforts taken—including mass vaccination—were nevertheless useful.

3. Methodology

This chapter provides a comprehensive overview of the methodology employed in the research, with a description of the DEA and the BoD methodologies, including the mathematical formulations inherent to the models. Additionally, it delves into the concept of CIs, offering a comprehensive understanding of their significance. The chapter proceeds to elucidate a specific modification pertaining to the implementation of CIs, that serves to enhance the model's faithfulness to real-world conditions, which refers to the integration of weight restrictions.

To better adapt the model to the case under analysis, the chapter elucidates the different types of returns to scale, delineating the situations in which they should be applied and highlighting their distinctions. To clarify how the model intended to assist healthcare decision-makers, the issue of the model's orientation is also up for discussion. Lastly, the chapter outlines diverse approaches to assess the quality of the model, fostering a comprehensive understanding of the research methodology.

3.1 Data Envelopment Analysis

Constructing a DEA model demands careful consideration of various aspects to ensure it aligns with the intended objectives (Jacobs et al., 2006). In subsequent subchapters, some key considerations will be highlighted, including a) Use CRS or variable returns to scale (VRS); b) Use input or output orientation; c) Weight restrictions; d) Dealing with slacks; e) Judging the quality of the model; f) How to adjust to environmental factors (Jacobs et al., 2006). Thus, in this chapter all these concepts will be developed in order to understand and choose the most appropriate model to evaluate the quality of care in Portuguese hospitals, providing a solid basis for subsequent analyzes, and greater reliability in the results obtained.

3.1.1 The concept

The DEA method was developed in 1978 by Abraham Charnes, William W. Cooper e Edwardo L. Rhodes, as part of Edward Rhodes' doctoral thesis at Carnegie Mellon University, supervised by Cooper (Charnes et al., 1995). The thesis's background is where the idea behind this methodology first came out, which aimed to assess the outcomes of follow-up programs for underprivileged students in American public schools, in the state of Texas. The proposal was to evaluate the impact on school achievement by comparing the effects of the program on schools that participated on the program and those that did not (Charnes et al., 1978). This approach was based on previous work carried out by Farrell (1957), who developed the production frontier analysis technique.

Over the years, DEA has found widespread application across various industries, including banking, insurance, agriculture, education, healthcare, transportation, and retail, among others. This versatile methodology has proven to be a valuable tool for decision-making, strategic planning, efficiency assessment, and identifying best practices in numerous sectors. DEA's introduction to the healthcare domain occurred in 1983 when Nunamaker utilized it to evaluate the efficiency of standard nursing services and, since then, it has played a crucial role in enhancing the operational efficiency of healthcare organizations and improving healthcare delivery.

DEA is a method based on non-parametric linear programming that enables to determine and compare the efficiency of multiple DMUs, which are organizational units responsible for converting multiple resources (inputs) into multiple results (outputs). These organizations must be homogeneous, a concept that is associated with the premise that all DMUs to be considered in the model must operate in the sense of performing identical tasks, as well as using the same type of inputs and, with these, producing the same type of outputs (Soares de Mello et al., 2003). Additionally, DEA is used to evaluate the best combinations of inputs and outputs, and to choose the appropriate weights to assign to each of these variables.

DEA emerges as a powerful methodology capable of addressing the inherent complexity in comparing units with multiple inputs and outputs, and yet has the ability to aggregate the results of these intricate comparisons into a single measure. This measure represents the efficiency of each analyzed DMU and is computed by considering the productivity index of the unit itself in relation to one or more reference units—those units with the best observed performance, referred to as benchmarks. Consequently, this method was developed to assess the efficiency of productive units without the necessity of considering financial aspects or converting inputs and outputs into monetary values, making it particularly valuable in contexts where such conversions are not relevant or desired. This eliminates the need for extensive financial data and complex adjustments, streamlining the efficiency assessment process.

Thus, DEA establishes an efficiency frontier by analyzing the best-performing observations, which are the extreme points. This approach utilizes mathematical programming techniques to effectively "envelop" or encompass the remaining data points within this frontier. Efficiency, in this context, is quantified as a ratio of the outputs produced to the inputs consumed, providing a numerical measure of how effectively a unit utilizes its resources to generate desired outcomes.

$$\text{Efficiency} = \frac{\text{Outputs}}{\text{Inputs}} \tag{1}$$

Therefore, an inefficient unit can enhance its efficiency through an expansion of products (output) while maintaining the same level of resources used, or through a reduction of resources (inputs) used while maintaining the same level of production, or through a combination of both strategies. This approach is best suited for non-linear input-output relationships and circumstances where typical performance measurements are either unavailable or challenging to calculate. It operates by maximizing the weighted output/input ratio for each provider, with the restriction that this ratio can equal, but never be greater than one, for any other provider in the data set (Charnes et al., 1978). Thus, the efficiency frontier reflects the highest degree of efficiency that any DMU in the data set is capable of achieving under best practice conditions. DEA makes it possible to compare units regardless of their variances in size, structure, and input/output mix.

In order to overcome the limitations of this method, researchers have proposed several ways to integrate it with other statistical techniques (Vara et al., 2023), to improve the robustness of the results obtained from DEA:

- Bootstrapping: it is a resampling method that enables to generate multiple samples from a single set of data, through a random selection of observations with replacement. Through this method, researchers can detect any outliers that may impact findings, thereby obtaining more accurate confidence intervals for DEA efficiency ratings (Nardo et al., 2005).

- Bayesian DEA: Bayesian DEA combines the DEA approach with Bayesian statistics, allowing the integration of prior knowledge and uncertainties regarding model parameters. Thus, researchers can obtain more accurate estimates of efficiency scores and predictions of future performance (Mitropoulos et al., 2015);
- Robust DEA: This methodology allows dealing with outliers and measurement errors, through statistical techniques, such as trimmed means and winsorization. This allows researchers to obtain more accurate efficiency scores even in the presence of extreme observations or data quality issues (Mitropoulos et al., 2015);
- Stochastic frontier analysis (SFA): SFA is a statistical approach that models the production process as a mixture of a deterministic component (the frontier) and a random error term. By combining these two methodologies (DEA and SFA), the stochastic nature of the production process can be considered, and more accurate efficiency ratings can be obtained (Vara et al., 2023).

3.1.2 Two stage DEA

As previously explained, DEA – also considered “first-stage DEA”- is a strategy used to measure the efficiency of DMUs, through the determination of efficiency scores for each DMU, defined in the interval]0,1]. However, several researchers have chosen to estimate a regression model using the results obtained in this stage to study the impact of exogenous factors on the efficiency of DMUs, a process that is known as "second-stage DEA". Thus, the method that involves both phases (“first-stage” and “second-stage”) is called “two-stage DEA” (Kontodimopoulos et al., 2006), and it is commonly used in production efficiency analysis to estimate the effects of operating conditions on performance. Therefore, the efficiency estimates acquired in the initial stage are combined with the contextual variables that indicate the operational setting in which the health systems are incorporated to create a regression of an econometric model. Thus, this analysis requires the presence of variables relating to sociodemographic factors and variables relating to the geographic distribution of the population (Costa, 2021). In this way, it is necessary to find out which external factors are commonly used in analyzes of this type, as well as which methods are most used in estimating the econometric model.

3.2 Benefit of Doubt

Motivated by DEA’s limitations, Cherchye et al. (2007) developed the BoD methodology, which is an extension of the original DEA model developed by Charnes et al. (1978) to the application in the construction of CIs. This methodology was originally proposed as a method of evaluating macroeconomic performance, however it has recently been adapted to the theory of indices.

This benchmarking approach does not focus on converting resources into products. In this method, inputs are not considered, using a dummy variable adjusted to one for all observations, and all sub-indicators are treated as outputs (Puyenbroeck, 2017). The dummy input may be easily comprehended by thinking of the model as a tool for aggregating a number of performance sub-indicators, without referring the inputs that are utilized to produce this performance. Thus, its outcome has a broader interpretation, being employed as a performance measure or estimate, which, depending on the variables used, is a broader concept than efficiency. In short, the BoD is a DEA with VRS, output-oriented and unitary inputs (Vara et al., 2023). These concepts will be further explained in sections 3.5, 3.6 and 3.7.

BoD methodology creates a CI for each DMU through the aggregation of different performance indicators into a single performance measure (Cherchye et al. 2007). To achieve this, a linear programming tool is utilized to improve the multipliers or weights linked to the indicators. For each entity, it is estimated a CI value equal to the maximum weighted arithmetic mean of the sub-indicators considered, with multipliers determined endogenously. Cherchye et al. (2007), the first author to implement this method, proposed obtaining the benchmarking unit by solving a maximization problem (considering that external benchmarks are also possible):

$$\mathrm{I}^* = \mathrm{I}^*(\mathrm{w}) = \operatorname{argmax} \left(\sum_{q=1}^{Q} \mathrm{I}_{qk}\, \mathrm{w}_{q}\right), \mathrm{k} = 1, \ldots, \mathrm{M} \tag{2}$$

I^* is the score of the hypothetical hospital that maximizes the overall performance (defined as the weighted average), given the (unknown) set of weights w. I_{qk} is the normalized score, applying the min-max method (which will be further explained in section 4.6), of the sub-indicator q^{th} (q=1,...,Q) for DMU k (k=1,...M), and w_q is the corresponding weight. Regarding this method, it is observed that:

(i) weights are hospital-specific: different sets of weights may lead to different hospitals being chosen, as long as none has the highest score on all sub-indicators;

(ii) the benchmark is generally hospital-dependent, so there is no single referral (unless, again, there is one better off across all sub-indicators);

(iii) the sub-indicators must be comparable, that is, have the same unit of measurement.

In order to represent the fact that the CI is a non-decreasing function of the indicators, multipliers are subject to a non-negative restriction, and a normalization requirement is also applied to the relative weighting. The indicator's final value cannot be greater than one if any other assessed entity applies the same set of weights (Karagiannis & Karagiannis, 2018). According to these authors, due to the fact that this methodology does not allow for negative indicators, this results in a limitation with regard to financial indicators.

The fundamental linear programming introduced by Charnes et al., (1978) can be expressed as follows, in which the weights are calculated endogenously for each DMU under evaluation:

$$\mathrm{CI}_{\mathrm{k}} = \max_{\mathrm{wik}} \sum_{i=1}^{m} \mathrm{w}_{\mathrm{ik}} \mathrm{y}_{\mathrm{ik}} \tag{3}$$

s.t.

$$\sum_{i=1}^{m} \mathrm{w}_{\mathrm{ik}} \mathrm{y}_{\mathrm{ik}} \leq 1 \quad \forall\ \mathrm{j} = 1, \ldots, \mathrm{n}, \mathrm{j} \neq \mathrm{n} \tag{4}$$

$$\mathrm{w}_{\mathrm{ik}} \geq 0\ \mathrm{i} = 1, \ldots, \mathrm{m} \tag{5}$$

CI_k is the value of the composite index for entity k, where the result will range between zero (worst possible performance) and 1 (the benchmark); w_{ik} is the multiplier (weight assigned) of sub-indicator i for observation k; y_{ik} is the value of sub-indicator i for DMU k; n is the number of observations under analysis; and m is the number of sub-indicators (Charnes et al., 1978). So, to calculate the CI for each DMU in this analysis, we must solve the program 37 times, one for each DMU j (Gaaloul & Khalfallah, 2014).

The constraints in this model guarantee that the multipliers that are calculated are positive and that no CI surpasses the unit value (Gaaloul & Khalfallah, 2014). Thus, Eq. (3) is the objective value to construct the CI. Eq. (4) is called the normalization restriction, which ensures that none of the CIs will be greater than one. In most BoD models, no normalization is needed, as the DEA can handle raw values (Cherchye et al., 2006). Eq. (5) guarantees the non-negativity of the computed weights. Since the model behind the calculation of the CIs is nonradial, the BoD-based composites cannot be viewed as factors imposing equal reduction/expansion of undesirable/desirable criteria, in a radial sense (Ferreira & Marques, 2021).

Equal weighting, opportunity-based weightings, and judgment-based weightings are common methods for selecting weights when a composite measure of healthcare quality is calculated from individual key performance indicators (KPI). The foundation for these strategies is the presumption that, at least in part, an organizational unit's stated choice over the relative relevance of indicators is reflected in relative performance on a set of indicators. As a result, a BoD strategy considers these preferences that have been disclosed by giving a higher weight to indicators with better performance, and lower weight to those with lower performance (Shwartz et al., 2009). This concept will be further explained in the next section.

According to this method, the indicators must be compensable, which refers to the possibility of making up with performance in one dimension for poor performance in other ones. This means that an entity from a hospital may outperform others in a certain dimension if it has an abnormally high value for an indicator to be maximized. As a result, it receives the highest performance rating, and all the other indicators are assigned with a null multiplier in that dimension (Calabria et al. 2016). This compensability feature enables a fair and thorough evaluation of performance by considering the trade-offs and advantages displayed by entities across several dimensions, offering a more accurate evaluation that accounts for the complexity of performance assessment in real-world contexts. However, some researchers argue that this approach can have problems, especially when there are indicators that cannot be easily offset by others. Therefore, in order to address these problems, other ways of calculating the performance index have been proposed, namely the "directional BoD". However, when applying the directional BoD to calculate the overall performance index, situations may arise where many indicators have null multipliers, meaning they do not contribute to the index calculation (Matos et al., 2021). To solve this problem, it is necessary to impose restrictions on the weights of the indicators, ensuring that all indicators contribute to the value of the CI. This concept will be further explained in section 3.4.

As stated by Carrilo and Jorge (2017), with this efficiency method it is possible to establish hospital classifications and rankings, which are becoming a more popular comparison tool, enabling the development of policies to improve the performance of providers.

3.3 Composite indicators

Efforts to improve quality measurement in healthcare involve an increasing number of quality indicators (QIs). While these QIs provide insights into specific aspects of healthcare quality, they often fail to capture all relevant data. To address this, combining these QIs into a composite score is useful for overall performance evaluation (OECD, 2008).

According to the OECD, a CI is created when many indicators are combined into a single index, based on a theoretical model of the assessed multidimensional concept. It measures multi-dimensional concepts that are impossible to express with a single indicator. Ideally, a CI should rely on a theoretical framework that guides the selection, combination, and weighting of individual indicators to accurately represent the subject's dimensions or structure (OECD, 2008). Thus, the CI reflects the overall performance of each DMU, incorporating variable weights, with values falling within the [0,1] range. Consequently, a CI value closer to 1 signifies a higher relative performance of the DMU compared to others.

Generic CIs are valuable when seeking a comprehensive view of DMUs. They offer a comprehensive perspective by encompassing all relevant categories, including financial and economic data. For instance, they can help pinpoint the areas where interventions are required to boost hospital performance (Morais & Camanho, 2011). These tools facilitate straightforward comparisons across units and can be applied to various domains, highlighting the intricacies of our dynamic environment (Gibari et al., 2019). In complex contexts, techniques like these may be incredibly helpful for directing decision-making and forming public opinion, since rather than presenting isolated or disaggregated data points, they simplify data analysis and communication. Furthermore, they enable the assessment of performance trends over time and the tracking of achieved outcomes (Pereira et al., 2021a). The OECD suggests a ten-step technique to construct a CI and describe it accurately (Nardo et al., 2008):

1. Theoretical Framework: This stage ensures that the concept is understood and defined clearly.
2. Data Selection: This phase calls for ensuring the accuracy of the indicators and data that have been chosen, identifying their advantages and disadvantages, and summarizing their key features.
3. Imputation of missing data: By estimating the values that are missing, it is crucial to present a full data set. The validity of the imputed values must also be provided.
4. Multivariate analysis: It should be utilized to examine the data set's general structure, judge its applicability, and direct further methodological decisions.
5. Normalization: This stage ensures that indicators may be compared while maintaining the theoretical framework and the characteristics of the data, by adjusting and/or transforming scales.

6. Weighting and aggregation: The underlying theoretical framework must be followed when performing this step. This ensures that the right weighting and aggregation of indicators are chosen in accordance with the theoretical framework and data attributes.
7. Uncertainty and sensitivity analysis: Must be carried out in order to evaluate the composite indicator's robustness in terms of indicator selection, normalization strategy, data imputed, and weights.
8. Back to the data: It is important to identify the key factors that contribute to overall performance, in order to maintain the credibility of the outcomes from earlier stages.
9. Links to other indicators: If the theoretical framework alone calls for it, this stage correlates the CI or its dimensions with other single or composite indicators that already exist.
10. Visualization of the results: This final stage is crucial since a proper depiction of the results reflects in a good interpretability. Therefore, it is crucial to make sure the visualization approach chosen is one that reveals more information in a more understandable way.

The vast diversity of CIs demonstrates their popularity as communication and policy assessment tools. However, these indicators remain debatable despite their growing recognition. Opponents of this tool use the absence of a standardized construction process and, in particular, the subjectivity inherent in its creation to cast doubt on its veracity (Chechye et al., 2007). In fact, subjective choices are disseminated in answering questions that are linked to the creation of a CI, such as: what general concept is intended to be summarized; which sub-indicators should be included in the analysis; how important should each sub-indicator be in the final calculation; how they should be aggregated; how to deal with missing or low-quality data; to what extent one can assess how DMU ratings are influenced by all of the above issues. Notably, CIs are controversial due to the potential for condensing provided information, which could potentially conceal crucial details.

These subjective assessments of the relative importance of each sub-indicator are incorporated through the attribution of weights, which signify the importance of the sub-indicator and give it value in relation to the others (Cherchye et al., 2007). These weights can either be fixed or variable, depending on the chosen approach. Several methods exist for establishing these weights, catering to the unique characteristics and preferences of different units. The most straightforward and commonly used strategy is equal weighting, where each sub-indicator is given equal importance. However, this approach may not accurately reflect the true significance of each sub-indicator, and it cannot address overlaps or correlations between indicators, potentially double-counting certain characteristics. Additionally, equal weighting assumes that all units should prioritize objectives in the same way, which may not align with the diverse goals and constraints faced by different entities. In healthcare, especially, it is challenging to apply fixed weights due to the many unknown variables and external factors influencing hospital efficiency (Zhou et al., 2007).

The second method involves deriving weights from perspectives, such as expert opinions or public consensus. Multi-criteria techniques like the Analytic Hierarchy Process (AHP) are often utilized for this purpose, which is a technique that allows both qualitative and quantitative aspects of an issue to be considered by eliciting pairwise comparisons between sub-indicators. However, interpreting these weights as significance coefficients can be misleading, as they represent trade-offs among sub-indicators rather than absolute importance. Since AHP relies on the comparability of participants' or experts' perspectives, it can introduce subjectivity and potential conflicts into the weighting process.

The third method for determining weights involves employing statistical techniques that aim to reduce subjectivity by deriving weights from sub-indicators in an endogenous manner. This approach includes methods like Principal Component Analysis (PCA), Factor Analysis, and deterministic techniques like DEA (Zhou et al., 2007).

3.4 Weight restrictions

Variable weights in DEA are determined individually for each hospital, optimizing its performance to its best potential regardless of the specific weight values assigned. This means that the performance is calculated to be maximized independently of the weight assignments, resulting in an input/output ratio that is optimized relative to that of all units. Thus, DEA allows a hospital situated on the efficiency frontier to assign higher weights to inputs and outputs in which it excels, while assigning very low weights to less significant factors. These variable weights, characterized by fewer constraints, often yield higher efficiency results compared to fixed-weight models. However, this complete freedom in weight selection can lead to secondary factors dominating the efficiency evaluation or critical factors being overlooked. To mitigate this, restrictions on weight allocation were considered to limit this selection freedom, which reduces the search region, preventing the efficiency of a DMU from increasing, and only allowing for potential decreases in comparison to unrestricted analysis.

To implement these weight restrictions, it becomes necessary to make value judgments concerning the relative significance of various outputs and the opportunity cost associated with the inputs utilized. Critics argue that such judgments may compromise the impartiality of unrestricted analyses. In response, Thompson et al. (1986) introduced a method for imposing these restrictions, known as the confidence region method. This approach involves constraining the ratios between the weights of inputs or outputs. Consequently, the confidence regions are based on the following relationships:

$$\alpha \leq \frac{u_r}{u_1} \leq \beta \tag{6}$$

And

$$\alpha' \leq \frac{v_i}{v_1} \leq \beta' \tag{7}$$

In this method, the lower (α, α') and upper (β, β') limits are established for the ratios between the weights of input and output variables, considering a reference variable (u_1 for output and v_1 for input) and the other variables (u_r for output and v_i for input, where r = 2,...,s and i = 2,...,m). When restrictions are included, the efficiency frontier undergoes changes (Thompson et al., 1986).

Despite requiring that the weights are not negative, as in Eq. (5), the BoD model allows them to be estimated freely, with the objective of maximizing the performance of the entity under evaluation, which also involves some disadvantages. In some cases, an entity may obtain a higher performance by assigning null weights to sub-indicators that present worse scores, implying that this sub-indicator associated with the null weight has no influence on the overall performance result. Therefore, this lack changes the CI scores, as they are now based on a lower set of sub-indicators. Since these sub-indicators are carefully selected for a credible analysis, it is important that they all intervene in the process of constructing the aggregate indicator. As such, to prevent this from happening, a lower limit is defined that requires the participation of all sub-indicators in the calculation of the CI, but keeping this lower limit free. In this way, weight restrictions serve as a valuable tool to enhance the model's discrimination when evaluating performance. However, it is important to acknowledge that this involves a value judgment, since it is the modeler who decides whether the situation will be better represented due to the implementation of such limits (Cherchye et al., 2007). Thus, the drawback of this strategy is that if there is an excessive restriction, the programming issue could not have a viable solution. As mentioned earlier, it is crucial to prevent the model from assigning zero weights to sub-indicators, therefore a lower limit will be defined for the sub-indicator's participation, keeping the upper limit free, as shown in the following Eq.:

$$\frac{w_{ik} y_{ik}}{\sum_{i=1}^{m} w_{ik} y_{ik}} \geq \alpha, \forall\ i = 1, \ldots, m \tag{8}$$

BoD approaches often face criticism for their limitations in enabling comparisons across different organizational units, as each unit employs distinct weightings. To illustrate this, consider a scenario where students are enrolled in a course with 5 distinct assignments, all of equal importance. There is a typical grading option that allows students to drop the assignment on which they have the lowest grade, before determining the final grade. In this case, the weighting mechanism is zero for the excluded task and a quarter for the other assignments. Another version of this strategy is to allow students to assign less weight to one of the tasks, such as only 12% weight on the task with the lowest score and 22% weight on the other four, which would decrease the likelihood that students would ignore 20% of the course material. In this way, only 20% of the class will be utilizing the same set of weights if the assignment is dropped (or downweighed) at random among students. There is no hesitation in determining a final grade based on the four (or five, if down-weighting is utilized) assignments and then determining a grade based on this score. As a result, it is generally understood that the comparison principle should be used when applying a common rule that may provide "self-serving" weights. This idea is generalized by DEA. As demonstrated by the large DEA literature, the fact that the rules result in various weights does not, in any way, affect the capacity to accurately compare performance across the units as long as all units "play under the same set of rules" (Shwartz et al., 2009).

3.5 Returns to scale

Returns to scale is an economic concept that pertains to the relationship between the size of production and the corresponding changes in inputs over time. In essence, return to scale theory examines how production responds to proportional increases in inputs when observed over the long term. This concept can be categorized into two primary types: VRS and CRS.

The first category encompasses both increasing returns to scale, where production grows at a higher rate than the change in inputs, and decreasing returns to scale, where production expands at a slower rate than the alteration in inputs. On the other hand, the second category pertains to situations in which production scales up in the same proportion as the change in inputs. For instance, if a 1-unit increase in input leads to a 0.5-unit increase in output, then a 10-unit increase in input would also result in a 5-unit increase in output. However, this assumption is not valid in many real-world scenarios, which is why alternative models that do not rely on constant returns to scale are often used.

3.6 Model orientation

Efficiency measurement models can be presented under two forms of orientation: input-orientation and output-orientation. The models that are oriented towards inputs are those that seek to optimize resource usage, which means using a lower quantity of inputs, while keeping the outputs constant. In other words, they intend to maintain production levels, but with a smaller number of resources. Conversely, an output-oriented model aims to maximize output levels using the available inputs while keeping their level fixed. For instance, in this type of orientation, the efficiency values obtained refer to how much the level of outputs can be increased proportionately while the level of inputs remains unchanged (Jia et al., 2017).

When choosing the direction to follow in a performance evaluation study, the decision maker should consider which variables he has more control over and which he wants to focus on. That is, they should choose an input-oriented model if they believe to have more control over the inputs, as they will have greater control over their reduction. On the other hand, if one feels to have more control over the outputs, then they should opt for an output-oriented model, as there will be a greater control over its increase. Assuming that managers have influence over both inputs and outputs, instead of giving priority to just one, a third type of model is known as a non-oriented, slack or additive model (Ozcan, 2008). In fact, the choice between the model orientations in the DEA does not affect the shape of the efficiency frontier, but rather the implications and decisions that may be made based on the obtained results.

3.7 CCR model vs BCC model

In their initial work, Charnes et al. (1978) introduced DEA with the assumption of CRS under the CCR model. This is the most basic model of DEA methodology, considered as the origin of DEA, and being the reference for the remaining models. This model includes CRS, and it is known as CCR in reference to the initials of the surnames of each of the authors who developed it (Charnes, Cooper and Rhodes).

Six years later, Banker, Charnes, and Cooper proposed an expansion of the original DEA model to address situations where not all DMUs are operating at an optimal scale, meaning that returns to scale are not constant. This variation in the relationship between inputs and outputs, as the scale of production changes, led to the development of the VRS model, also known as BCC, in reference to its authors, Banker, Charnes, and Cooper. This model allows considering potential economies of scale, accommodating variations in productivity that occur when the scale of production is adjusted. Thus, this approach is particularly valuable when different units of analysis operate at varying production scales, and may experience efficiency gains or losses as they adjust their scale of operation. In essence, the researcher assesses efficiency under various returns to scale and determines whether the observed levels align with the efficiency frontier, corresponding to a particular return to scale. So, instead of having the assumption of proportionality between inputs and outputs, the BCC model has the assumption of convexity. In this model, each unit is assessed in relation to other units that are closer in size to it, creating a frontier that further limits the range of possible positive results. As a result, a value is produced referring to the measurement of technical efficiency (Oliveira & Ferreira, 2014).

As can be seen in the following graph, the CCR model is considered a more demanding model, since by incorporating CRS it implies that efficient DMUs are those that operate at the optimal point of scale, meaning that any deviation from this scale results in inefficiency. In turn, the BCC model considers an efficiency cone around each efficient DMU, formed by the crossed planes that involve the data represented by points, allowing greater flexibility in the evaluation of efficiency. Due to this difference, it is common to observe that efficiency results obtained in the BCC model are greater than or equal to those obtained in the CCR model.

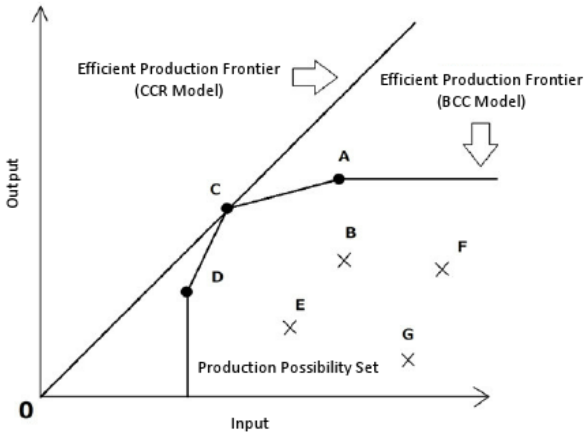

Figure 3- The CCR model and the BCC model production possibility set (Source: Xiaoya & Jinchuan, 2008).

The efficient frontier of the CCR model is the dotted line that passes through C from the origin. The frontiers of the BCC model consist of the bold fines connecting A, C and D. The production possibility set is the area consisting of the frontier together with observed or possible activities with an excess of input and/or a shortfall in output compared with the frontiers. Thus, A, C and D are on the frontiers and so they are BCC-efficient. The same happens for all points on the solid lines connecting A and C, and C and D. However, only C is CCR-efficient (Xiaoya & Jinchuan, 2008).

The decision to use either CRS or VRS is pivotal, as operating at a suboptimal scale in healthcare often arises due to external factors. This choice depends on the specific context and objectives of the analysis, as well as whether it pertains to a short-term, productivity-focused perspective or a long-term, managerial-oriented one (Jacobs et al., 2006). Therefore, opting for a CRS model is more suitable when adopting a productivity-centric perspective, whereas a VRS model is better suited for a managerial perspective.

3.8 Model quality

DEA has the advantage of not requiring assumptions regarding the production frontier, thus not being necessary to specify a particular production function with fixed parameters. Instead, this method uses a nonparametric approach to measure the relative efficiency of DMUs by comparing their inputs and outputs. Due to this absence of a production function, it is impossible to carry out statistical tests that assess the quality of the model. Nevertheless, it is crucial to scrutinize whether the chosen model effectively and robustly represents reality. Therefore, when examining the model, special attention should be given to the aspects outlined in the following sections.

3.8.1 Adjust to environmental factors

For a study that uses the DEA methodology to evaluate the activity of the units under analysis, it is also necessary to control the aspects extrinsic to them, that is, those particularities that the decision-makers do not control. Each DMU implies certain characteristics or factors that imperatively influence its performance (Castro, 2011). Thus, to account for this influence, one method is to estimate the efficiency frontier using the conventional DEA method in the initial stage of the analysis and then, based on the results, create a regression model using environmental variables, in which the performance results of DMUs are used as a dependent variable (Barbetta et al., 2007). However, another way to control these impacts was proposed by Jacobs et al. (2006), and consists of three ways of considering environmental factors in efficiency analysis:

1- Limit the comparison to organizations that operate in similarly limited environments.
2- Model the restrictions as being comparable to the elements in the process of production;
3- Carry out risk adjustment.

The first approach attempts to resolve the issue of numerous environmental constraints faced by different organizations, by grouping DMUs into clusters that share similar environmental characteristics. In this type of organization, a more accurate comparison of efficiency between organizations is created, as they are analyzed within more homogeneous subsets and in comparable environments. However, the disadvantage is that it becomes a challenge to distinguish between fluctuations in performance caused by external influences, and differences in efficiency. Furthermore, due to the exclusion of organizations with multiple environmental factors from investigation, this method may also lead to a reduced sample size.

In turn, the second strategy involves adding environmental variables as inputs directly into the production model, considering their impact on productivity, treating them as components of the industrial process (Burgess and Wilson, 1995). In this way, the DMUs will be compared with others that operate in worse or similar environments, being considered more efficient than those that operate in the worst environments.

Lastly, when performing risk adjustment, the results are modified under various conditions before being incorporated into the efficiency model. This strategy offers a more specialized analysis, as it allows you to adjust each result based on specific characteristics related to the product, rather than using environmental variables as a general adjustment for all results.

3.8.2 Comparability between the considered DMUs

In performance analysis, it is generally assumed that all DMUs operate in a similar way. However, their specific contexts can significantly affect performance, especially in a scenario like this study, where DMUs are hospitals with varying resources, patient demands, and regional characteristics. Different specialties and sizes among hospitals bring unique qualities and challenges. For example, a cardiology-focused hospital may have higher expenses due to specialized needs, while a pediatric hospital may incur costs for child-friendly infrastructure and trained staff. Additionally, a hospital's geographic location can affect efficiency, as someone in an area with a prevalence of chronic diseases may face more difficult challenges in managing resources and caring for patients. Therefore, analyzing the CIs for each DMU is vital to understanding the factors that contribute to performance and the contextual factors that affect efficiency.

3.8.3 Error measurements

In any real-world data collection, there is the potential for measurement error. This means that measurements may not be 100% accurate due to a variety of factors, such as inaccurate measuring instruments, human errors in data collection, or natural fluctuations in measurements. The data used for analysis in this study are taken from open databases, which contain missing values for many DMus and time periods, as well as discrepant data that do not make sense in the context where they are found. As such, when interpreting results, it is crucial to be aware of this potential source of error and consider how it may affect the analysis conclusions.

3.8.4 Outlier observations

An observation that differs abnormally from other values in a population-based random sample is called an outlier. As explained previously, one of the disadvantages of the DEA methodology is the fact that, due to the lack of prior knowledge of the production function under study, it is more sensitive to erroneous data and outliers (Vasquez, 2012), so it is crucial to address this sensitivity when conducting the analysis. Any statistical study must begin with the precise quantification and determination of all variables. However, the real world often introduces uncertainty, unpredictability, and data gaps, making these inherent challenges (French, 1995).

In this analysis, the values that appeared as outliers for each indicator, which means the values that were discrepant, were analyzed. Values greater than 100% were replaced by the average value of the referring indicator, as well as null values of variables that would not make sense to present such results. To handle missing data, mean imputation was employed, which is a method that involves assigning missing values the average value of the corresponding indicator for the remaining DMUs. Despite being an easy method to implement, it does have the disadvantage of potentially attenuating correlations involving the imputed variable(s), making it less suitable for multivariate analyses (Ferreira et al., 2023). Missing data can occur randomly or non-randomly. However, sometimes it is challenging to determine its nature, and most imputation methods assume a random fault mechanism. If there is reason to believe that missing values follow a non-random pattern, it should be explicitly modeled and incorporated into the analysis.

3.8.5 Inclusion and exclusion of variables

The selection of performance indicators to be integrated into the model should accurately represent the activities and objectives of the analyzed DMUs, offering a comprehensive reflection of their performance across relevant domains. However, acquiring such detailed and exhaustive information can be challenging, primarily due to restricted data accessibility. Any inclusion or exclusion of a sample from the model will inevitably lead to alterations in the results. Furthermore, when considering the inclusion or exclusion of variables, it is important to note that the greater the number of variables considered, the lower the discriminatory power of the model, potentially resulting in more hospitals being classified as top performing. To address this issue, specific rules have been established to govern the relationship between the number of DMUs used and the number of variables to be incorporated into the model, as will be elaborated in section 4.5.

Maintaining consistent variable scales is crucial to avoid distorting performance outcomes, so it is essential to ensure that all variables fall within the same order of magnitude. Variables expressed as percentages can be included in the analysis only if the entire set of variables adheres to this format, which is the case of this study.

3.8.6 Sensitivity and uncertainty analysis

The combination of two analyses, uncertainty analysis (UA) and sensitivity analysis (SA), allows a deeper assessment of the robustness of the CI, increasing its transparency. UA focuses on how the uncertainty inherent to the inputs propagates through the structure of CI, affecting its values. In turn, the SA focuses on how much each source of uncertainty contributes to the variance of the outputs (Saisana et al., 2005). Thus, it is a test for the existence of sensitive variables in the model, helping to identify which are the indicators or weights whose inclusion/exclusion has a more significant influence on the final results. Such SAs are not restricted to varying the components of given observations, but also to eliminating some observations entirely, and/or introducing additional observations. In this context, this analysis helps to understand how changes in indicators' weights may impact the values of CIs (Saisana et al., 2005).

In the field of constructing CIs, these two methods are normally used separately. However, their combined application proved to be more powerful (Nardo et al, 2005). According to this author, it is essential to consider all the possible sources of uncertainty related to the construction of a CI, namely: a) the model chosen to estimate the measurement error in the data, e.g., based on the information on variance estimates that is currently accessible; b) the mechanism for including/excluding sub-indicators from the CI; c) the transformation and/or cutting of sub-indicators, such as removing outliers; d) the type of normalization strategy used to eliminate scale effects from the sub-indicators, e.g. re-scaling or standardization; e) the amount of missing data and the imputation procedure selected to replace missing data; f) the choice of the weights, such as equal weights or weights obtained using a DEA-based method; g) the level of aggregation, if more than one level is employed, e.g., at the indicator or at the sub-indices level; h) the choice of the aggregation system, e.g., additive, multiplicative, multi-criteria analysis, etc.

Saltelli et al. (2008) propose a method to gain insights into the impact of various sources of uncertainty by conducting simulations. Instead of altering the indicators themselves, one can modify factors such as the data normalization methods, the weights assigned to them (which are randomly generated), or the aggregation techniques. This approach aids in comprehending the contributions of these different aspects to overall uncertainty.

4. Case study: the Portuguese public hospitals

In this section, we delve into a comprehensive case study designed to perform a benchmark analysis of the quality of care provided by Portuguese public hospitals. We will begin by shedding light on the intricacies of the Portuguese NHS and the impactful reforms that have shaped the Portuguese healthcare sector, providing an essential context for our study.

Subsequently, the sources of information that support this analysis will be referenced, detailing the origins of the data. This will be followed by an explanation of the sample selected for the study, including its composition, and the specific period under analysis. After that, the data and variables that constitute the basis of our study will be referenced. An overview of health performance indicators will be provided, differentiating between desirable and undesirable outcomes. In addition, a concise summary of the variables incorporated into the model will be made, offering clarity regarding their composition and significance.

To ensure a comprehensive assessment, environmental factors that have the potential to impact the performance of healthcare institutions will be considered. Finally, the specificities of the model that will be used for the benchmarking analysis will be described.

4.1 The Portuguese national health system

Prior to the development of the Portugal's NHS, the 19th and 20th centuries witnessed significant influences from politics and religion on the organization of healthcare services. These influences led to gradual adjustments, in response to emerging health challenges. During that period, families, private institutions, and social security's medical and health services belonged to the State, which was in charge of assisting the underprivileged people [3].

In 1979, the NHS was officially established through Law No. 56/79, enacted on September 15th. This public and universal health system is based on the principles of generosity, universality, and solidarity. The Portuguese Constitution characterizes this system as "being universal, providing globally integrated care, being free of charge, guaranteeing equity, and having regionalized organization and decentralized management", with the primary objective to provide all Portuguese citizens with immediate and equitable access to high-quality medical care. It is based on the Beveridge system, wherein survival and financial independence are guaranteed through public funds and social contributions, through taxes paid by the general public. This system is governed and overseen by the Ministry of Health, which creates health policies, specifies clinical guidelines, and monitors the quality of services provided (Pereira et al., 2021b).

The NHS consists of a network of healthcare services serving the entire nation, from primary care to specialist and hospital treatment. The starting point of contact between patients and the healthcare system is hospital centers, which offer primary care. In addition to these, this system comprises university hospitals, public hospitals, and other specialized healthcare facilities. These hospitals provide a comprehensive variety of medical, surgical, and diagnostic services to address the population's most

[3] SNS. (2023). Portal do SNS: https://www.sns.gov.pt/sns/servico-nacional-de-saude/, accessed on March 14, 2023.

complicated health requirements. Thus, the health care provided by the NHS is ensured by several entities, namely:

- Groupings of health centers (ACES, abbreviation for the Portuguese words “Agrupamentos de centros de saúde”): Organizations responsible for the delivery of primary healthcare to regional populations. Each ACES covers a number of medical facilities and serve as the NHS's fundamental units for providing non-urgent circumstances with support and health care. The National Vaccination Program's immunizations are all administered there, along with standard medical, nursing, and family planning consultations. Usually, a municipality's area is covered by the hospital center;
- Local health units (ULS, abbreviation for the Portuguese words “Unidades locais de saúde”): Comprise all clinics and hospitals in a city or area into a single, integrated facility that is in charge of offering primary and specialty healthcare;
- Hospitals: Usually, these units specialize in providing advanced medical care. Most of them function as integrated hospital centers, overseeing multiple hospitals within the same city or region.

A fundamental aspect of this healthcare system is its dedication to evidence-based medicine and the promotion of impartiality in access to healthcare. In this way, the NHS employs a triage and prioritization system that is based on the severity and needs of patients, aiming to provide equal access to healthcare for all citizens. This approach ensures that patients with the most urgent medical conditions receive priority treatment, thereby reducing waiting times, and ensuring immediate and efficient care. While the NHS serves as the primary healthcare provider in Portugal, the private sector also complements the public system, offering individuals the option to access private healthcare services (Nunes & Ferreira, 2018).

From 1990 onwards, a series of issues became evident in the Portuguese NHS, including inefficient resource allocation, low productivity, lengthy waiting times, dissatisfaction among users and healthcare professionals, and high healthcare service costs. Several methods were adopted to solve these problems, but they proved to be inefficient. A pivotal moment in the evolution of Portugal's healthcare system occurred in 1993 with the approval of the new status for the NHS. This transformation included a major modification, namely the establishment of five health regions throughout the country (Nunes & Ferreira, 2022a):

- North Health Region: headquartered in Porto, whose area of influence covers the districts of Braga, Bragança, Porto, Viana do Castelo, and Vila Real;
- Central Health Region: headquartered in Coimbra, whose area of influence covers the districts of Aveiro, Castelo Branco, Coimbra, Guarda, Leira, and Viseu;
- Lisbon and Tagus Valley Health Region: based in Lisbon, whose area of influence covers the districts of Lisbon, Santarém, and Setúbal;
- Alentejo Health Region: based in Évora, whose area of influence involved the districts of Évora, Beja, and Portalegre;
- Algarve Health Region: based in Faro, whose area of influence included the district of Faro.

With consideration for the distinctive characteristics and requirements of each region in the country, these regions were created with the intention of decentralizing the management and provision of healthcare. This strategy made it possible for health services to be located closer to local communities, enabling more region-specific decision-making and the implementation of healthcare policies tailored to local conditions. Presently, each health region is overseen by a Regional Health Administration (ARS, abbreviation for the Portuguese words “Administração Regional de Saúde”), which assumes responsibility for coordinating and managing healthcare services at the regional level (Nunes & Ferreira, 2022a). The ARS possesses legal status, administrative and financial independence, as well as its own assets. One of the primary functions of health regions is to assess the performance of institutions and healthcare services, alongside their roles in planning, resource allocation, guidance, and coordination of activities [4].

In 2002, a new hospital management model was introduced, granting greater autonomy to management bodies, and leading to significant changes in the Basic Health Law. Subsequently, in 2003, the primary health care network was created, which aims to provide health care to the citizen in permanent articulation with hospital and continuous health care, which allowed for a more integrated and effective approach in the provision of care for citizens. In 2005, there was a significant transformation in the management of public hospitals, shifting them from public limited companies (SA, derived from the Portuguese abbreviation "Sociedade Anónima”) to EPE, granting them the status of state-owned public companies. This change provided hospitals with greater flexibility and resources to enhance the efficiency of resource utilization within the NHS. Recognizing the challenges posed by an aging population, increased life expectancy, and a higher prevalence of chronic illnesses, the National Network of Integrated Continuous Care was established in 2006. This network's mission is to improve people's autonomy and quality of life by offering integrated care services at various complexity levels.

In 2007, the first family health units were introduced as part of the primary healthcare reform. These units were designed to focus on prevention, diagnosis, and personalized healthcare, providing comprehensive health services to families. Subsequently, in 2008, the Hospital Center Groups of the NHS were created to establish a robust and balanced organizational framework for delivering primary healthcare services. This initiative aimed to enhance the effectiveness of healthcare management, and patient-centered access to services.

In 2011, the adoption of electronic prescriptions marked a significant step towards digitizing the healthcare system. This innovation facilitated communication among healthcare professionals from different organizations, reducing the likelihood of errors or misunderstandings in medication prescriptions, and aiding in the fight against fraudulent practices. In the same manner, in 2012, the introduction of the User Portal allowed citizens to register and access online services, such as appointment booking and procedure verification, enhancing convenience and accessibility for healthcare services. In 2017, the Health Research Fund was finally established to support research initiatives aimed at promoting, protecting, and improving people's health.

[4] SNS. (2023). Portal do SNS: https://www.sns.gov.pt/sns/servico-nacional-de-saude/, accessed on March 22, 2023.

The most recent and pivotal event within the scope of the NHS pertains to the administration of COVID-19 vaccines. This measure was implemented in response to the ongoing global COVID-19 pandemic, which originated in 2019, with Portugal recording its first case of infection in 2020. The impact of this pandemic reverberated throughout healthcare systems worldwide, including the NHS. This health system assumed a crucial role in responding to the public health crisis, managing the spread of the virus, treating the infected population, and implementing preventive measures to mitigate the pandemic's impacts. Given the potential for significant fluctuations in statistical outcomes based on the year in which this event occurred, it represents a critical timeframe to be considered in this research.

4.2 Reforms in the Portuguese health sector

Since its inception, the Portuguese NHS has undergone through several reforms. These transformations have been driven by global concerns, including an aging population, the need to strike a balance between cost containment and ensuring access and quality of care, as well as a shift from acute to chronic illnesses. Notably, Portugal, alongside other OECD countries, has implemented substantial changes in its healthcare system over the past decade (Khushalani & Ozcan, 2017). One standout initiative in Portugal has been the restructuring of the hospital system. This endeavor initiated a process of care integration, encompassing both vertical and horizontal integration, resulting in the consolidation of numerous public hospitals. Consequently, this restructuring gave rise to hospital facilities and ULS (Ferreira et al., 2018b).

In Portugal, recent administrative reforms have seen the emergence of paradigms such as governance and the new public service. These paradigms have pushed the managerial reform perspective to consider the input of various social actors throughout the reform process, from its definition to implementation and subsequent adjustments. However, it is important to note that the New Public Management (NPM) paradigm has had the most significant impact on administrative reform in Portugal over the last three decades, leading to a redistribution of wealth and social protection in the country, a development unprecedented before (Madureira, 2020). Starting from the mid-1990s, marked by innovative management practices, and later in 2002, influenced by NPM principles, the political agenda for healthcare in Portugal has also incorporated extensive reforms within the NHS. These reforms have been oriented towards meeting the needs of users and fostering greater efficiency and quality in the management of healthcare services, particularly in the hospital sector (Nunes & Ferreira, 2018; Nunes & Matos, 2017). In this way, according to Nunes (2020), the adoption of the NPM perspective in the health sector promoted several points, namely debureaucratization and decentralization, through the implementation of business management; a user-centered approach; greater transparency; accountability and results-oriented management; demand for high performance; development of market mechanisms; stimulation of competition between providers; cooperation with new operators; introduction of contracting; and improvement of the quality of health services.

4.3 Source of information

The data utilized in this research were sourced from the official ACSS website[5], which is the governmental body responsible for overseeing and managing the healthcare system in Portugal. Established in 2007, it operates as a public institute integrated into the indirect administration of the state, equipped with administrative and financial autonomy. ACSS plays a pivotal role in executing the directives set forth by the Ministry of Health, while also being subject to its supervision. Thus, it is noteworthy that ACSS exercises its jurisdiction over the entire continental territory of Portugal (Vara et al., 2023). Moreover, data from the PORDATA website[6], a reputable platform known for providing comprehensive and reliable statistical information, was also incorporated into the study, for a more comprehensive analysis. Additionally, certain data related to the hospital units under examination were sourced from their respective Account Reports.

4.4 Sample and period of analysis

According to the INE, as of 2021, Portugal was home to a total of 240 hospitals. This figure reflects a slight decline of one hospital compared to the preceding year but marks a notable increase of 11 hospitals since 2010. Among these hospitals, there were 128 privately-owned hospitals, signifying an increase of 26 facilities since 2010, 110 public hospitals, and two establishments operating as PPP hospitals. Within the category of public hospitals, 105 were designated as universal access hospitals, while the remaining 5 were either military or prison hospitals.[7]

When selecting the DMUs to be analyzed in this study, three key factors were considered: the sample size, which represents the number of DMUs; the ratio of DMUs to the number of model variables (performance indicators); and the homogeneity of the production units. It is worth noting that, on the one hand, it is desirable to incorporate as many DMUs as possible, since a larger sample enhances the likelihood of capturing high-performing units that can determine the efficient frontier, thereby improving discriminatory power. On the other hand, an excessively large dataset can cause the homogeneity of this dataset to decrease, which would mean that some exogenous impacts could affect the results (Sarkis, 2007). Additionally, it might reduce the methodology's ability to distinguish between different units effectively. Therefore, considering the importance of ensuring the homogeneity of the chosen sample, it is essential to highlight its potential limitations. Incorporating heterogeneous DMUs can introduce bias into the results and complicate their interpretation, thus the DMUs included in the study might exhibit comparability in terms of size, operations, available resources, and other relevant characteristics. To ensure the homogeneity of production and structural processes, a selection process was conducted, resulting in the exclusion of certain institutions from the final set of DMUs.

As a part of the selection process, a list of DMUs was compiled, each of which had data available for the previously chosen performance indicators. Initially, data were collected for all public hospitals with information accessible on the ACSS website, resulting in a list of 55 hospitals. Subsequently, a refinement process was undertaken, excluding certain DMUs that did not possess data for all the

---

[5] BENCHMARKING ACSS: http://benchmarking.acss.min-saude.pt/, accessed on April 5, 2023.
[6] PORDATA: https://www.pordata.pt/, accessed on April 8, 2023.
[7] INE (Instituto Nacional de Estatística): https://www.ine.pt, accessed on May 13, 2023.

performance indicators. In order to enhance comparability and bolster the homogeneity of the sample, the following categories of DMUs were excluded: maternity wards, psychiatric hospitals, psychiatric recovery centers, hospitals resulting from PPPs, Portuguese oncology institutes, regional alcoholology centers, as well as orthopedic and ophthalmology units. Additionally, municipalities in the Madeira and Azores islands were excluded due to substantial data gaps. This meticulous curation process led to a reduced sample of 38 DMUs. Furthermore, during the analytical phase, it was observed that one DMU, "Hospital Vila Franca de Xira", had a significant amount of missing data, particularly for the year 2019. This extensive absence of data raised concerns about the accuracy of the results, ultimately resulting in the removal of this DMU from the sample. This stringent approach ensured that the final list exclusively comprised units with comprehensive and dependable data, thereby mitigating the risk of source bias (Ferreira et al., 2018b). Consequently, the final sample encompasses 37 Portuguese healthcare institutions operating as EPE. The study's analytical timeframe was established to cover the period from 2019 to 2021, resulting in a sample of 111 entries (37x3=111). The comprehensive list of DMUs selected for inclusion in the analysis can be found in Table 1.

Table 1- Sample used in analysis (Source: Author).

Abbreviations	Hospitals	Abbreviations	Hospitals
CHBM	Centro Hospitalar Barreiro/Montijo	CHULN	Centro Hospitalar Universitário Lisboa Norte
CHL	Centro Hospitalar de Leiria	CHVNGE	Centro Hospitalar Vila Nova de Gaia/Espinho
CHLO	Centro Hospitalar de Lisboa Ocidental	HSOG	Hospital da Senhora da Oliveira Guimarães
CHS	Centro Hospitalar de Setúbal	HB	Hospital de Braga
CHBV	Centro Hospitalar do Baixo Vouga	HDFF	Hospital Distrital da Figueira da Foz
CHMA	Centro Hospitalar do Médio Ave	HDS	Hospital Distrital de Santarém
CHO	Centro Hospitalar do Oeste	HESE	Hospital Espírito Santo de Évora
CHUC	Centro Hospitalar e Universitário de Coimbra	HPDFF	Hospital Professor Doutor Fernando Fonseca
CHEDV	Centro Hospitalar Entre Douro e Vouga	HGO	Hospital Garcia de Orta
CHMT	Centro Hospitalar Médio Tejo	HSMM	Hospital Santa Maria Maior
CHPVVC	Centro Hospitalar Póvoa de Varzim/Vila do Conde	ULSG	Unidade Local de Saúde da Guarda
CHTS	Centro Hospitalar Tâmega e Sousa	ULSCB	Unidade Local de Saúde de Castelo Branco
CHTV	Centro Hospitalar Tondela-Viseu	ULSM	Unidade Local de Saúde de Matosinhos
CHTMAD	Centro Hospitalar Trás-os-Montes e Alto Douro	ULSAM	Unidade Local de Saúde do Alto Minho
CHUCB	Centro Hospitalar Universitário Cova da Beira	ULSBA	Unidade Local de Saúde do Baixo Alentejo
CHULC	Centro Hospitalar Universitário de Lisboa Central	ULSLA	Unidade Local de Saúde do Litoral Alentejano
CHUSJ	Centro Hospitalar Universitário de São João	ULSN	Unidade Local de Saúde do Nordeste
CHUA	Centro Hospitalar Universitário do Algarve	ULSNA	Unidade Local de Saúde do Norte Alentejano
CHUP	Centro Hospitalar Universitário do Porto		

Each of these entities plays a crucial role in delivering healthcare to the Portuguese population, being divided into 3 categories: Hospital Center (CH, abbreviation for the Portuguese words "Centro Hospitalar”), Hospital, and ULS. To provide a clearer understanding of these concepts, the definition of each one will be presented, according to INE:

- Hospital Center - A public health entity with administrative and financial autonomy, possessing its own assets, that integrates several hospitals lacking legal personality. These hospitals share common services and functional connections.
- Hospital - A healthcare establishment that offers curative and rehabilitative services in both inpatient and outpatient settings. Hospitals may also engage in disease prevention, teaching, and scientific research.

- Local Health Unit (ULS) - A public health entity composed of hospitals/hospital centers and ACES. These units ensure the integrated provision of primary, hospital, and continuous healthcare to the population residing within their respective areas of influence.

4.5 Data and variables definition

This section will detail the performance indicators used to assess hospital quality within the scope of this study. It will also elucidate the criteria used in the selection of these indicators and the logic behind their importance in this research.

An indicator must demonstrate relevance, based on its influence on health, significance for policy, and propensity to be impacted by the healthcare system. Thus, its evaluation must be practical, and the information required to generate it must be readily available, trustworthy, while balancing the reporting burden and substantiating scientific validity (Mattke et al., 2006). In the context of this study, the chosen indicators must possess the capacity to compare institutions and evaluate important aspects of healthcare quality. When choosing the variables that will be included in the model, a topic of concern and discussion is involved - the relationship between the number of DMUs and the number of variables that measure performance - when applied the DEA methodology. Therefore, when confronted with a scenario where the number of DMUs is limited and the indicators are abundant, the discriminatory power and precision of this methodology in assessing each DMU's performance diminish. Consequently, it becomes imperative to maintain a higher number of DMUs in relation to the number of variables, as emphasized by Khezrimotlagh et al. (2019). Several researchers have addressed this question, and notable guidance comes from the influential work of Golany and Roll (1989). Their research suggests a rule of thumb regarding the relationship between the number of DMUs, inputs, and outputs, noting that it is advisable that the number of DMUs exceeds the sum of the number of variables (inputs and products), multiplied by 2. In simpler terms, there should be more than 2 DMUs for each input and each product, and this principle can be expressed mathematically as follows:

- Nº DMUs > 2 x (Nº inputs + Nº outputs)

On the other hand, other researchers, such as Cooper et al. (2007), have put forward a related guideline. According to their approach, the number of DMUs should exceed the sum of the number of inputs and outputs, but this time multiplied by 3:

- Nº DMUs > 3 x (Nº inputs + Nº outputs)

In turn, Dyson et al. (2001) proposed a slightly different criterion, suggesting that the number of DMUs should surpass the product of the number of inputs and the number of outputs multiplied by 2:

- Nº DMUs > 2 x Nº inputs x Nº outputs

As per Bogetoft and Otto (2012), linear programming theory stipulates that if an optimal solution exists, then there is an optimal basic solution where the number of positive variables is at most equal to the number of linear constraints. Denote m as the number of inputs, n as the number of outputs, and K as the number of DMUs. In the BCC model, there are m+n+1 constraints and K+1 variables (λ and θ). Since θ is typically positive, there are a total of $m + n + 1 - 1 = m + n$ variables that must be positive. In the CCR model, which has one less constraint, the number of variables should be either m+n or m+n−1.

The text suggests that adding too many variables can undermine the analysis's robustness, making it hard to differentiate top and underperforming DMUs. In this study, ten variables were chosen for 37 DMUs, adhering to the mentioned criteria.

4.5.1 Healthcare performance indicators

To determine the most popular and widely accepted indicators in studies that assess hospital quality, a literature review was previously carried out, followed by a consultation of the official NHS benchmarking database. This encompasses a wide array of indicators pertaining to healthcare in Portuguese hospitals, meticulously categorized across various domains. Therefore, the choice of variables considered several key factors, such as the existing literature, data availability within the chosen sample and the time frame, the quality of data, and their relevance to the study's objectives. Give that the BoD methodology will be applied in this benchmarking analysis, primacy was given to relative variables, as they allow a more balanced and fair evaluation of the DMUs, considering the uncertainty in the data. As this approach treats all sub-indicators as outputs and does not consider the input side (Morais & Camanho, 2011), the selected indicators were categorized into desirable and undesirable outcomes. As such, the most employed variables in the literature are presented in the graphs below, usually divided into inputs and outputs.

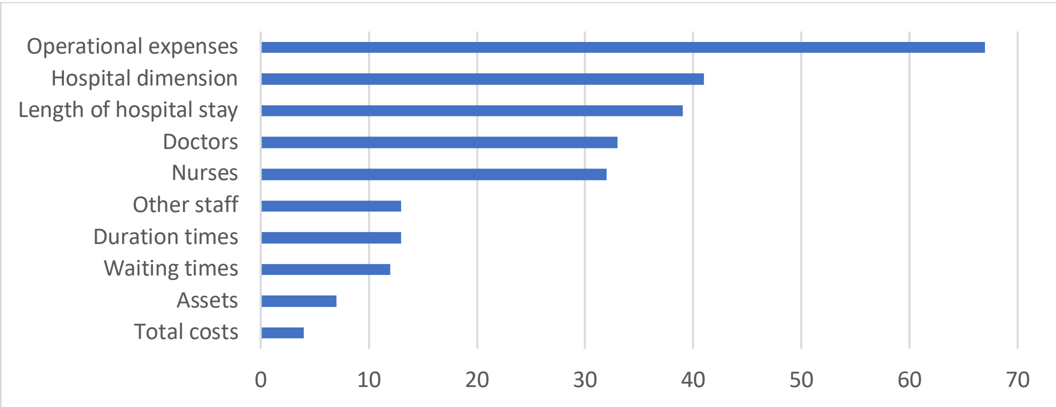

Figure 4- Most used inputs in reviewed studies in the healthcare area (Source: Author).

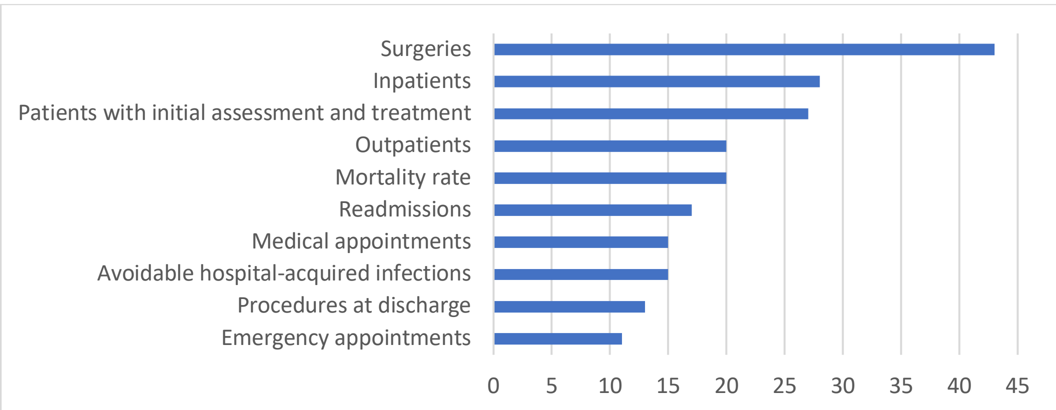

Figure 5- Most used outputs in reviewed studies in the healthcare area (Source: Author).

The analysis of these graphs reveals that certain variables are notably more abundant compared to others, with operational expenses, surgeries, hospital dimension, length of hospital stay, and patients with initial assessment and treatment standing out in terms of quantity. However, the most frequently used variables may not always be the most suitable for defining the quality in healthcare, as many of the studies examined aimed to compare institutions in terms of efficiency or other aspects. Therefore, within this group of variables, which encompass various indicators, the ones that most accurately captured the hospital's quality were considered.

To verify the interrelationships among these variables and detect any redundancy, a Pearson correlation analysis was conducted, as detailed by Ferreira et al. (2019). This analysis ensures that each of the remaining variables contributes with new and non-redundant data to the model. The Pearson coefficient assumes that the dataset exhibits monotonous variation, a proportional (linear) trend, and a normal distribution (Miot, 2018). It is worth noting that, despite some variables deviating from a normal distribution, as indicated by a p-value below 0.05 in the Shapiro-Wilk test, conducted using IBM SPSS Statistics 25 software (see Tables 9, 10 and 11 in the Appendix), the Pearson coefficient was still employed (See table 12 in the Appendix). This decision was influenced by the substantial sample size, consisting of 370 observations. Although the presence of outliers can potentially impact the direction and magnitude of this coefficient, it was deemed that outliers have a more substantial impact in smaller samples. In light of this, the next sections will reveal the list of the ten chosen performance indicators.

4.5.1.1 Desirable outcomes

(q1) Occupancy rate- Represents the percentage relationship between the total number of days of hospitalization throughout the year and the unit's capacity. It is a relevant measure of access, as it is closely related to the waiting time and the availability of resources (beds) (Aloh et al., 2020). In line with the findings of Lechintan[8], the ideal occupancy rate typically falls within the range of 85% to 90%. Nevertheless, there are multiple healthcare institutions where this indicator exceeds the recommended threshold. According to Marahem et al. (2018), an occupancy rate surpassing 90% can precipitate the risk of collapse and overcrowding, a situation that may compel hospitals to decline patient admissions and/or delay healthcare provision. Conversely, when the value of this indicator falls below 85%, it signifies inefficient and unequal resource allocation, thereby posing risks. As such, the optimal value defined for this indicator was 85%.

(q2) Percentage of external consultations with discharge record- Represents the percentage of outpatient visits that culminated in patient discharge. Usually referred to as "ambulatory visits" or simply "outpatients", this metric covers consultations carried out on an outpatient basis, where the patient is examined and subsequently discharged from the hospital. This indicator suggests that these patients received the necessary care, without the need for new referrals or hospitalization. As such, a higher value for this indicator is viewed positively, as it demonstrates that the hospital is efficiently managing the needs of outpatients, without requiring additional resources or prolonged treatment, improving patient flow, and maximizing resources utilization.

[8] Lechintan, A. (2017). April's smartKPI: % Hospital Bed Occupancy Rate. https://www.performancemagazine.org/smartkpi-hospital-bed-occupancy-rate/, accessed on July 21, 2023.

This indicator is a commonly used outcome measure in a substantial portion of the research documented in the literature. In fact, it ranks among the top five most frequently employed outcome measures in the articles examined for this study. Several authors have employed this variable in similar analyses, including Magnussen (1996), Linna et al. (2006), Nayar et al. (2013), Portela et al. (2016), and Ferreira and Marques (2021).

(q3) Percentage of emergency episodes attended within the expected time- Represents the proportion of emergency episodes that were attended to within the predefined timeframe. In the context of hospital benchmarking, evaluating the performance of healthcare units in emergency care is of paramount importance since this metric directly impacts patient satisfaction and clinical outcomes. To emphasize the importance of this indicator, Nunes (2017) highlights that “the payment of emergency services to the NHS hospitals that provide them will take into account the percentage of episodes treated within the waiting time provided for in the triage protocol”.[9] This measure was implemented with the primary objective of reducing waiting times in the emergency department, as extended wait times can have adverse effects on both the patient health and the reputation of the hospital. Therefore, various researchers have incorporated this indicator into their performance assessments, with a common focus on strategies aimed at reducing emergency room wait times, to enhance organizational performance (Backman et al., 2016). Additionally, several authors also have used this indicator as a means of evaluating the performance of their DMUs, including Grosskopf et al. (2004), Garcia-Lacalle and Martin (2010), Barros et al. (2007), Afonso and Fernandes (2008), Barros et al. (2013), Portela et al. (2016), Ferreira et al. (2018a), and Pereira et al. (2021b).

(q4) Percentage of first appointments performed in appropriate time- Represents the proportion of first consultations that were carried out within the appropriate time based on the patient’s condition, directly influencing service quality. There exists a legislatively established maximum guaranteed response time (TMRG, an abbreviation for the Portuguese words "tempo máximo de resposta garantido") for initial consultations in hospitals following appointment requests (as defined in Decree-Law No. 44/2017). Consequently, a higher percentage in this indicator signifies superior overall performance. In this way, this variable serves as a gauge of whether the hospital is effectively providing timely healthcare services to the community. As revealed in the literature review, timeliness plays a pivotal role in ensuring good evolution of a community's health indicators, including increased life expectancy and reduced mortality rates, since timely medical interventions can prevent potential complications.

As of December 31, according to the ERS (in Portuguese “Entidade Reguladora da Saúde”), there were 581,909 individuals awaiting their first specialty consultations in an NHS hospital, following a referral from a primary care physician. Alarmingly, nearly half (47%) of these individuals were waiting for consultations that exceeded the maximum allowable time frame. This situation serves as a clear

[9] Nunes, Q. (2017). Tempo de espera nas urgências vai contar no pagamento da tutela aos hospitais: https://www.linkedin.com/pulse/tempo-de-espera-nas-urgências-vai-contar-pagamento-da-quental-nunes/?originalSubdomain=pt, accessed on August 3, 2023.

indication of subpar healthcare service quality and may also reflect patient dissatisfaction. It is crucial to note that delays can have adverse effects on patients' health, exacerbate existing conditions, and lead to frustration and discontent. Given this, the direction of this indicator will be to maximize.

(q5) Percentage of patients operated within the TMRG (P1 180 days)- Represents the percentage of patients who undergo surgical procedures within the TMRG, which, in this context, is set at 180 days. Based on the extensive literature review conducted for this research, the most commonly employed performance indicator in articles related to healthcare efficiency and performance analysis is "surgeries." Within the dimension of access, this indicator encompasses various subcategories, including "surgeries canceled," "scheduled surgeries," "hip fracture surgeries," and more. The TMRGs are defined by Ordinance No. 1529/2008, dated December 26th. TMRG for surgeries spans from 72 hours (Priority Level 4) to 270 days (Standard Priority), following the surgical recommendation. Urgent procedures require rapid intervention and are not factored into this scenario.

This indicator has substantial relevance in such analyzes for several compelling reasons. Firstly, it has a direct link with the quality of surgical services, given that a high value of this indicator suggests that the hospital can meet scheduling commitments, guaranteeing quick access to necessary surgical procedures. Furthermore, this metric reflects the hospital's management capacity with regard to planning and organizing surgeries, and may indicate that the hospital has effective procedures to plan and execute surgeries within the prescribed deadlines, facilitating patient flow management, and optimizing the use of resources.

(q6) Percentage of surgical outpatient services (surgical interventions)- Represents the percentage of surgeries conducted on an outpatient basis, signifying cases in which patients are admitted and discharged on the same day, without needing hospitalization. Ambulatory surgery represents a valuable tool for enhancing patient care, effectiveness, and hospital efficiency. It not only allows for the allocation of hospitalizations to more complex cases, but also rationalizes healthcare expenses. As defined by the Directorate-General for Health, outpatient surgery is defined as a scheduled surgical intervention performed under general, locoregional, or local anesthesia. These procedures can take place in dedicated facilities but are typically conducted within an inpatient setting. Importantly, they are performed safely and in accordance with contemporary medical standards, with patients admitted and discharged on the same day, obviating the need for hospitalization. Ambulatory surgery has witnessed global growth, contributing to an increase in the efficiency and quality of care, concurrently reducing costs for the NHS. As a result, it is delivering a healthcare service that places a stronger emphasis on patients and their requirements, thereby fostering heightened accessibility to hospital care (Mota, 2016).

4.5.1.2 Undesirable outcomes

(q7) Percentage of hospitalizations with a delay of more than 30 Days- Represents the proportion of hospitalizations that exceed the period of 30 days concerning the total number of hospitalizations with discharge in the analyzed timeframe, which is a duration exceeding the deemed appropriate period. In the majority of studies, a closely related variable, often referred as "length of stay," was commonly utilized. Several authors, including Afonso and Fernandes (2008), employed this as a measure of resource consumption, a proxy for healthcare demand, and an efficiency indicator. As highlighted by OECD, "All other things being equal, a shorter stay will reduce the cost per discharge and shift care from inpatient to less expensive post-acute settings."[10] Therefore, a high value for this indicator may indicate inefficiencies in the healthcare system, where patients remain in hospitals longer than necessary, resulting in increased costs, and a shortage of available beds for other patients. Longer hospital stays can also be associated with a higher risk of infection and a decline in treatment quality. Thus, shorter hospital stays may hinder adequate patient follow-up care and potentially lead to adverse health outcomes. Zhan and Miller's study (2003) also suggested that shorter stays are associated with higher-quality care.

(q8) Percentage of expenses with provision of services in total expenses with personnel- Refers to the proportion of expenses allocated to the provision of services in relation to the total expenses with personnel. According to the literature, the most frequently employed indicator in performance-related studies in the healthcare is “operating expenses”, which encompasses various categories of expenditures. Within this indicator, it was found that the most widely studied costs involve expenses related to service provision, medical supplies, and expenses with personnel remuneration. This indicator allows us to examine the allocation of financial resources to staff recruitment and compensation, compared to those allocated to direct patient care. Thus, it provides insights into resource distribution, and the balance between personnel investment and service accessibility. A lower proportion in this indicator may indicate more efficient utilization of personnel-related resources, with a smaller expense share going to service provision. This suggests a prudent resource allocation, reducing waste and ensuring optimal personnel utilization. Essentially, it serves as an indicator of effective management, enabling hospitals to deliver quality services with fewer financial resources allocated to staff.

(q9) Percentage of readmissions in 31-180 Days- Refers the proportion of hospital discharges, which marks the formal release of a patient, that subsequently resulted in readmission, within a window spanning from 31 to 180 days following the initial discharge. The significance of readmissions as a performance metric is multifaceted. It holds importance due to the value it offers to patients (Balla et al, 2008), the financial implications it carries (Ashton et al., 1995) and, most notably, it is closely tied to the concept of quality within healthcare. The link between hospital readmissions and healthcare quality is rooted in two key assumptions. Firstly, it implies that patients receiving high-quality care are typically discharged when they are clinically stable, reducing the need for readmission. Secondly, premature discharge or suboptimal initial care can lead to post-discharge complications, increasing the likelihood of readmission (Thomas, 1996). Therefore, unplanned readmissions may indicate inadequately addressing patients' medical needs. While the ideal time frame for readmissions varies, many authors use the 30-day cutoff, such as Thomas and Holloway (1991), Halfon et al. (2006), Ferreira and Marques (2018), Ferreira et al. (2020), and Matos (2020).

[10] OECD (2023). Length of hospital stay (indicator). https://data.oecd.org/healthcare/length-of-hospital-stay.htm, accessed on August 14, 2023.

(q10) Pressure ulcer rate- Refers to the proportion of patients who develop pressure ulcers in relation to the total number of patients under observation or care. According to the literature, it is frequently expressed as the number of pressure ulcer cases per 1,000 or 10,000 patients. Also known as bedsores or decubitus wounds, pressure ulcers are localized areas of necrosis that affect regions subjected to pressure. That is, these are injuries that occur in the skin and underlying tissues due to continuous pressure on certain areas of the body, primarily affecting individuals with limited mobility, or those confined to extended periods of bed rest. Therefore, the primary purpose of this indicator is to monitor the quality of care provided, particularly in long-term care facilities. In this sense, a reduced incidence of pressure ulcers reflects a better approach in preventing and managing these injuries, leading to improved health outcomes for patients. Notably, this indicator ranks among the most frequently employed metrics in the literature in performance analyses within the healthcare field. It has been featured in studies conducted by various authors, including Laine et al. (2005a), Shwartz et al. (2009), Backman et al. (2016), Ferreira and Marques (2021), and Pereira et al. (2021a).

4.5.2 Environmental factors

Environmental factors, often referred to as exogenous variables, are introduced to explore whether uncontrollable indicators can influence the quality of healthcare providers. Among these factors, sociodemographic variables play a prominent role and are extensively utilized in such analyses.

Sociodemographic factors encompass characteristics or attributes of a population that provide insights about its the social and demographic composition. They serve as key indicators for classifying and scrutinizing various facets of a population's characteristics, behaviors, and trends. Additionally, they help elucidate the social and economic structure of a specific group of individuals. Consequently, these factors are categorized as uncontrollable or exogenous variables, and have substantial importance in areas such as public health, social sciences, market research, political analysis, and more. In the context of this analysis, sociodemographic factors assume primary importance, providing a broader context for interpreting the quality of healthcare organizations in relation to the characteristics of the populations residing in the respective hospital region (Costa, 2021).

Derived from the extensive literature review undertaken in this research, the ten exogenous variables most commonly featured in the analyzed articles are as follows: ownership status, hospital bed capacity (as an indicator of hospital size), age distribution of the resident population, inhabitants (total population), geographical location (urban or rural), region, teaching status, educational attainment and literacy rates, purchasing power, mortality rates, and system characteristics (affiliation, centralization, decentralization, or independence). This data is graphically depicted in the accompanying chart for better visualization.

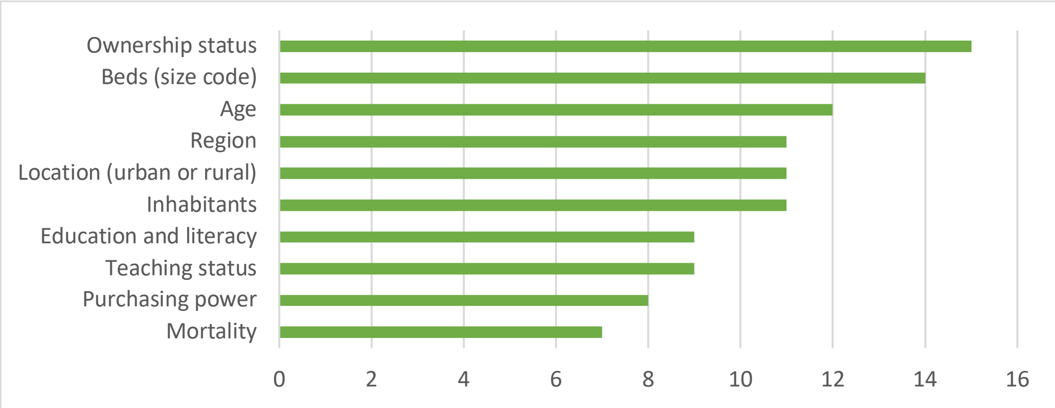

Figure 6- Most used sociodemographic variables in reviewed studies in the healthcare area (Source: Author).

The selection process of exogenous variables was guided by considerations such as insights from the literature, their relevance in the domain of hospital quality, availability of data for analytical purposes, and their propensity to exhibit minimal correlation with other exogenous variables, as it is not desirable for such correlation to occur between independent variables in the same analytical model. For this study, where all healthcare entities in the sample share the same ownership status (public, belonging to the NHS), variables such as "system characteristics" and "ownership status" will not be integrated into the analysis, as they are unlikely to offer discriminatory insights. Furthermore, the "purchasing power" variable was also omitted due to lack of data from 2020 and 2021.

Comprehensive statistical data were collected for each of the selected variables in all municipalities in the country where the hospitals in the study sample are located. Subsequently, an in-depth analysis of the interrelationships between these variables was carried out, which involved the use of statistical methods, to assess the strength and nature of these relationships. This analysis will be explained in greater detail in section 5.4. As a result, eight exogenous variables were carefully selected, which will be described below, accompanied by explanations about their interpretation, and justifications for their selection within the scope of this study.

(e1) Population density - This indicator reflects the average number of individuals per square kilometer (km²) in a given area. The population density of a region can significantly impact the demand for healthcare services, subsequently affecting the workload and resource distribution within hospitals.

(e2) Aging index - This variable represents the ratio of elderly individuals (those aged 65 and above) to young individuals (those aged 0 to 14) in a given municipality, expressed as a percentage. A higher percentage of elderly individuals within the population can signify more diverse and intricate healthcare requirements, as older individuals often necessitate specialized medical attention. Additionally, the elderly population tends to experience a higher prevalence of health issues, which can complicate surgical procedures, extend patient care durations, and increase the demand for hospital services.

(e3) Beds in hospitals - This indicator represents the quantity of inpatient beds available in hospital facilities. It indicates the hospital's size and reflects its ability to accommodate the local population's healthcare needs.

(e4) Teaching Status - This is a binary indicator, assigned a value of one for teaching hospitals, which are hospitals affiliated with educational institutions or universities, where educational and training activities for healthcare professionals occur. Conversely, it is assigned a value of zero for hospitals lacking such educational affiliations. Teaching hospitals often benefit from advanced resources and highly skilled staff.

(e5) Education - This variable indicates the proportion of the population that has attained a secondary education level, equivalent to completing up to the 12th grade of schooling. Given that a basic level of education is often necessary to comprehend health-related matters, this factor may be correlated with the ability to make informed decisions regarding medical care. In this way, it enables to analyze whether more educated population exerts an influence on the quality of the health system.

(e6) Mortality - Refers to the number of deaths in a given municipality per 1000 residents. The region's mortality rate serves as an indicator of the overall health status of the local population, as regions with elevated mortality rates may face increased demand for healthcare and hospital services.

(e7) Patients per Doctor - This variable measures the average number of inhabitants per physician within a specific municipality, serving as an indicator of the accessibility of healthcare services. An uneven distribution of doctors can lead to a lack of availability of health professionals in a given region. Indeed, residents may face difficulties in accessing timely healthcare services, potentially resulting in delays for essential medical procedures, which can have a cascading effect.

(e8) Doctors Per Inhabitants - This variable quantifies the number of doctors per 1000 residents within a specific municipality. Therefore, it serves as a measure of doctor availability within an area in relation to the resident population, which directly reflects the accessibility of healthcare services in that region.

Data collection for each of these variables was obtained through information available on the official INE website[11] and in the PORDATA[12] database. Given that the data is segmented by municipality and certain entities span multiple municipalities - exemplified by institutions like Centro Hospitalar Barreiro/Montijo and Centro Hospitalar Póvoa de Varzim/Vila do Conde, EPE - a weighted average was calculated, utilizing the population distribution as the determining weights. Once all variables were established, the following multiple linear regression model was formalized, in order to obtain an answer to the second research question (R_2):

$$\begin{aligned} \mathrm{CI} = \beta 0 + \beta 1 * \mathrm{PopulationDensity} + \beta 2 * \mathrm{AgingIndex} + \beta 3 * \mathrm{Beds} + \beta 4 * \mathrm{TeachingStatus} + \beta 5 \\ * \mathrm{Education} + \beta 6 * \mathrm{Mortality} + \beta 7 * \mathrm{ResidentsperDoctor} + \beta 8 * \mathrm{DoctorsperInhabitants} \\ + \mathrm{u} \end{aligned}$$

β are the parameters that will be estimated in the regression and μ is the error term. To estimate the regression, *IBM SPSS Statistics 25* software will be used.

[11] INE (Instituto Nacional de Estatística): https://www.ine.pt, accessed on September 2, 2023.
[12] PORDATA: https://www.pordata.pt/ , accessed on September 2, 2023.

4.6 Model specifications

Utilizing the BoD model outlined in section 3.2, which incorporates both desirable and undesirable indicators, a CI will be formulated to represent the quality of each entity. As already mentioned, in this method the indicators are typically assumed to have a positive contribution to the hospital's performance due to their non-negative values (Caldas & Varela, 2023). Nevertheless, there are instances where this assumption does not hold true. To account for this, we need to standardize or scale the indicators, as recommended by (Cherchye et al., 2007, 2011):

$$l_{ij}^t = \begin{cases} \dfrac{\mathrm{max}l_i - l_{ij}^t}{\mathrm{max}l_i - \mathrm{min}l_i}, \text{if the indicator is undesirable} \\ \dfrac{l_{ij}^t - \mathrm{min}l_i}{\mathrm{max}l_i - \mathrm{min}l_i}, \text{if the indicator is desirable} \\ 1 - \dfrac{|o_i - l_{ij}^t|}{\mathrm{max}l_i - \mathrm{min}l_i}, \text{otherwise} \end{cases} \tag{9}$$

Let l_{ij}^t denote the observation associated with the i_{th} indicator (i = 1, ..., m), the j_{th} hospital (j = 1, ..., n) and the moment t. In Eq. (9), o_i represents the hypothetical optimal value associated with indicators that are neither desirable nor undesirable. There are some cases where indicators have a value (or range) in which the ideal indicator should reside, therefore they should not be maximized or minimized. This is an example of the variable q1 (occupancy rate), which, as previously mentioned, must vary between 80% and 90%. Therefore, we must define o_i as the midpoint of the interval, that is, (90 + 80)/2 = 85%, which will be considered the optimal value (Caldas & Varela, 2023). After applying Eq. (9), the resulting indicators will be scaled to a range between zero and one, in order to make them comparable in space and time, being clear that the higher the better (Vidoli et al., 2023).

Thus, to assign the direction of improvement for each indicator in relation to its ideal value, the min-max normalization method was employed, as outlined in Eq. (9), an approach that allows to assign polarities—whether positive or negative—to each indicator. This is a method that can be used in conjunction with all weighting schemes, such as the BoD, and for all aggregation systems (OECD, 2008). Therefore, indicators that ought to be minimized, meaning that lower values indicate better performance, were assigned negative polarity and, conversely, indicators that should be maximized, signifying improved performance with higher values, were assigned positive polarity. The exception is the variable "occupancy rate", whose optimal value is 85%, obtaining higher values when the indicator value approaches this value.

To this end, there must be a set of m non-negative weights that have the function of maximizing the CI for hospital j in instant t, so that no other set of weights can result in a higher CI value. These weights, as previously mentioned, are also called multipliers, and are unit and time dependent. Since all indicators must be considered for the construction of the CI, the weights must not obtain null values, but above a limit, $\zeta > 0$, denoting the minimum acceptable value for a weight: $w_{ij}^t \geq \zeta, i = 1, \ldots, m, j = 1, \ldots, n$, which in this analysis was assumed to be $\zeta = 0.05$, as in the work carried out by Caldas and Varela (2023).

5. Results and discussion

This section presents the analysis and interpretation of the results obtained, in order to achieve and explain the defined objectives of this study. For the first phase of the analysis, the CIs were meticulously computed using the *R Studio* software, leveraging the *Compind* package. This package encompasses a range of functions for constructing CIs, including a function for the BoD model and the BoD model with weight restrictions for the sub-indicators, which were utilized in this research. Furthermore, the analysis carried out to address the second research question outlined in this work was conducted using *IBM SPSS Statistics 25* software, highlighting the influence of external variables on the values obtained in the previous question.

5.1 Variables synthesis

In this section, a concise table will be presented with the ten variables used in this case study, accompanied by their respective descriptive statistics. To compile the final data presented in the table, descriptive statistics were initially collected for each variable monthly throughout the year. However, during the analysis of this monthly data, numerous outliers and anomalous values were observed, some of which exceeded values of 200%, and higher. To resolve this issue and ensure the robustness of the results, we chose to use annual data. Therefore, the average values of each indicator were previously calculated for each DMU, using the monthly values present in the databases. Finally, using the *R studio* software, annual descriptive statistics of the variables were obtained for each year of the analysis, mitigating the impact of monthly fluctuations, and presenting more consistent and representative results.

Table 2- Descriptive statistics of the variables (Source: Author).

		2019	**2020**	**2021**
Occupancy rate	Mean [1)]	75,93%	76,43%	80,30%
	AAVR [2)]		0,50%	3,87%
	Minimum [3)]	68,72%	60,13%	63,43%
	Maximum [4)]	98,22%	92,02%	96,13%
	Median [5)]	87,13%	76,81%	80,32%
	Standard Deviation [6)]	6,48%	6,30%	6,75%
Percentage of external consultations with discharge record	Mean	17,04%	14,56%	16,01%
	AAVR		-2,48%	1,45%
	Minimum	4,28%	4,07%	3,07%
	Maximum	46,39%	42,88%	69,09%
	Median	16,47%	13,88%	14,35%
	Standard Deviation	7,93%	7,39%	10,75%
Percentage of emergency episodes attended within the expected time	Mean	73,58%	79,94%	76,79%
	AAVR		6,36%	-3,15%
	Minimum	22,46%	66,06%	99,61%
	Maximum	99,78%	99,56%	99,61%
	Median	73,58%	78,70%	76,79%
	Standard Deviation	13,58%	7,84%	8,84%
Percentage of first appointments performed in appropriate time	Mean	70,74%	60,44%	77,67%
	AAVR		-10,30%	17,23%
	Minimum	49,19%	40,84%	54,02%
	Maximum	91,83%	81,96%	97,46%
	Median	71,75%	62,04%	77,65%
	Standard Deviation	10,91%	10,96%	10,77%

Percentage of patients operated within the TMRG (P1 180 days)	Mean	80,48%	77,59%	82,41%
	AAVR		-2,89%	4,82%
	Minimum	55,43%	56,44%	61,95%
	Maximum	98,03%	93,45%	98,49%
	Median	81,05%	78,39%	84,17%
	Standard Deviation	10,31%	9,60%	9,34%
Percentage of surgical outpatient services (surgical interventions)	Mean	66,27%	63,55%	67,91%
	AAVR		-2,72%	4,36%
	Minimum	47,86%	41,38%	47,85%
	Maximum	78,59%	80,65%	85,20%
	Median	66,37%	63,01%	68,13%
	Standard Deviation	6,67%	8,55%	8,42%
Percentage of hospitalizations with a delay of more than 30 Days	Mean	3,85%	4,32%	4,15%
	AAVR		0,47%	-0,17%
	Minimum	1,04%	1,81%	1,73%
	Maximum	5,97%	7,23%	8,16%
	Median	3,72%	4,26%	4,05%
	Standard Deviation	1,15%	1,20%	1,26%
Percentage of expenses with provision of services in total expenses with personnel	Mean	6,59%	6,12%	6,48%
	AAVR		-0,48%	0,36%
	Minimum	0,84%	0,59%	0,89%
	Maximum	17,19%	17,92%	20,16%
	Median	6,10%	5,49%	6,13%
	Standard Deviation	4,23%	4,08%	4,38%
Percentage of readmissions in 31-180 Days	Mean	10,02%	9,23%	8,69%
	AAVR		-0,79%	-0,54%
	Minimum	4,37%	6,50%	6,00%
	Maximum	13,61%	12,63%	11,96%
	Median	10,06%	9,19%	8,80%
	Standard Deviation	1,60%	1,33%	1,24%
Pressure Ulcer Rate	Mean	0,05%	0,09%	0,11%
	AAVR		0,04%	0,02%
	Minimum	0,01%	0,01%	0,02%
	Maximum	0,16%	0,29%	0,29%
	Median	0,06%	0,07%	0,09%
	Standard Deviation	0,04%	0,08%	0,07%

1) Since the data for the variables are presented on a monthly basis, we first calculated the monthly averages and subsequently computed the annual average for the entire dataset in each year.
2) The Average Annual Variation Rate (AAVR) was calculated as the evolution of the average value of the variable in the year in question, compared to the previous year.
3) The minimum refers to the sample average value of the minimum monthly values of the total sample.
4) The maximum refers to the sample average value of the maximum monthly values of the total sample.
5) The median refers to the sample mean value of the monthly median values of the total sample.
6) The standard deviation was calculated based on the monthly average values of the total sample.

Upon a brief examination of the results, it becomes apparent that the average occupancy rate exhibited an upward trend over the three years of the study. This escalation is unsurprising, given the context of the ongoing pandemic, as outlined in section 4.1, which began to significantly impact health in Portugal from 2020 onward. Notably, there was a substantial 10-percentage-point decline in the average percentage of first appointments performed within an appropriate timeframe in that year, followed by a significant recovery the following year. Concerning the remaining variables, there were no significant increases or decreases in average values; rather, minor fluctuations within the study period were observed. Therefore, a more detailed analysis of these indicators will be provided in the subsequent section, with a focus on each of the hospitals within the sampled dataset. Moreover, the small standard deviations observed in the table indicate low heterogeneity within the sample studied. This implies that the values of the variables tend to be consistent and show little variation over the years, suggesting that the health units analyzed present relatively uniform results in relation to the metrics considered.

5.2 Portuguese hospitals performance

In this particular case study, there is not a constant reference entity that remains the same throughout the analysis period. Instead, the reference entity tends to change over time, and in some instances, even on a monthly basis. Therefore, rather than selecting a single reference entity, an analysis was conducted to identify the primary entities, and how frequently they were regarded as benchmarks over the course of the 12 months, in each of the three years examined (2019, 2020, and 2021). For example, if the two main entities are considered benchmarks for 12 of the 36 months analyzed, this would equate to a percentage of 30%. This percentage indicates that when choosing the two primary benchmark entities, one of these two entities served as a benchmark for approximately one-third of the months under review. The frequency with which each entity was categorized as a reference during the 36-month observation period is shown in table 14 in the Appendix.

Examining the table, it becomes evident that the five hospitals most frequently designated as benchmarks during the study period were CHPVVC (52,8%), HDFF (44,4%), HSMM (44,4%), CHEDV (38,9%), and CHMA (38,9%). Nonetheless, it is important to acknowledge that this percentage does not account for the values of the CIs during the remaining months, i.e., the months when they were not designated as a benchmark. Therefore, to gain a comprehensive view of their performance over time, it is also necessary to analyze the average results of the CIs they attained annually over the period under analysis, as a hospital might have been a benchmark in one month but performed poorly in the subsequent month. Therefore, to analyze quality over time, the average value for the CI that the units obtained in each year was considered, as represented in the following graph.

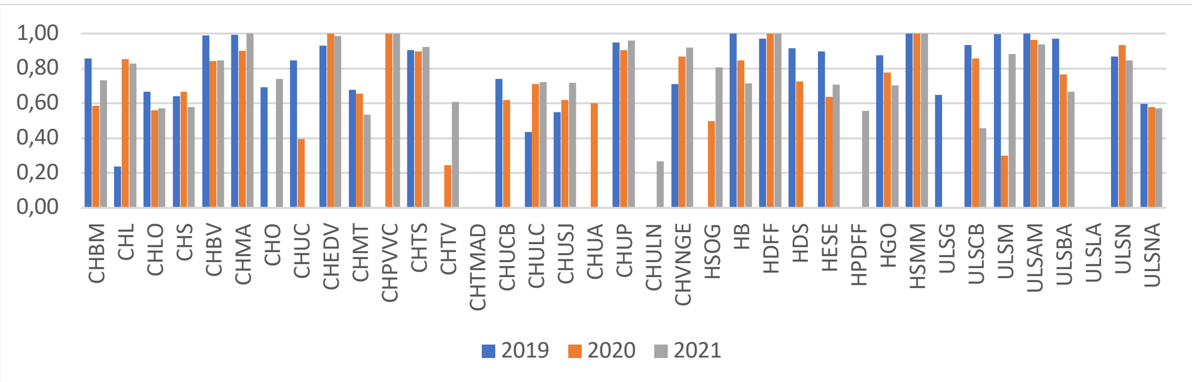

Figure 7- Evolution of the value of the CIs for each unit, over the three years under analysis (Source: Author).

According to the data obtained, only one healthcare entity consistently maintained its benchmark status over the entire three-year analysis period, which was Hospital Santa Maria Maior (HSMM). Although some of the hospitals classified as benchmarks changed during this period, the count of hospitals with an annual CI of one remained at four. Therefore, the hospitals that could be categorized as reference institutions—those with an index value exceeding 0.9—were subjected to a more comprehensive analysis. Table 3 presents the results obtained for each CI over the three years, along with their average values to ascertain the final reference status.

Table 3- Final CIs of the quality (Source: Author).

	2019	2020	2021	Global CI
CHBM	0,86	0,58	0,73	0,73
CHL	0,24	0,85	0,83	0,64
CHLO	0,67	0,56	0,57	0,60
CHS	0,64	0,67	0,58	0,63
CHBV	0,99	0,84	0,85	0,89
CHMA	0,99	0,90	1,00	0,96
CHO	0,69	0,00	0,74	0,48
CHUC	0,85	0,39	0,00	0,41
CHEDV	0,93	1,00	0,98	0,97
CHMT	0,68	0,65	0,53	0,62
CHPVVC	0,00	1,00	1,00	0,67
CHTS	0,91	0,90	0,92	0,91
CHTV	0,00	0,25	0,61	0,28
CHTMAD	0,00	0,00	0,00	0,00
CHUCB	0,74	0,62	0,00	0,45
CHULC	0,44	0,71	0,72	0,62
CHUSJ	0,55	0,62	0,72	0,63
CHUA	0,00	0,60	0,00	0,20
CHUP	0,95	0,91	0,96	0,94
CHULN	0,00	0,00	0,27	0,09
CHVNGE	0,71	0,87	0,92	0,83
HSOG	0,00	0,50	0,81	0,43
HB	1,00	0,84	0,71	0,85
HDFF	0,97	1,00	1,00	0,99
HDS	0,92	0,72	0,00	0,55
HESE	0,90	0,64	0,71	0,75
HPDFF	0,00	0,00	0,55	0,19
HGO	0,88	0,78	0,70	0,78
HSMM	1,00	1,00	1,00	1,00
ULSG	0,65	0,00	0,00	0,22
ULSCB	0,94	0,86	0,46	0,75
ULSM	1,00	0,30	0,88	0,73
ULSAM	1,00	0,96	0,94	0,97
ULSBA	0,97	0,76	0,67	0,80
ULSLA	0,00	0,00	0,00	0,00
ULSN	0,87	0,93	0,85	0,88
ULSNA	0,60	0,58	0,57	0,58

Between 2019 and 2021, the number of hospitals classified as reference institutions amounted to 13 (35%), 8 (22%), and 9 (24%) hospitals, respectively. It is worth highlighting ULSM, which exhibited a performance decline from 2019 to 2020 (-70%), which was then significantly recovered in 2021, resulting in a 66% increase. In contrast, HDS experienced a substantial performance drop, with an index of 0.92 in 2019, declining by 22% in 2020, and reaching zero in 2021, marking a 100% decrease. HB also displayed a 16% performance decrease from 2019 to 2020, followed by a 15% decrease in the subsequent year. HESE experienced a decline over the analyzed period, with a 29% drop from 2019 to 2020 and a 9.8% decrease in the following year. The CHBV unit likewise observed a 15% decrease in performance between 2019 and 2020, maintaining a similar figure in the subsequent year. Other reference hospitals in 2019 also registered performance declines, although not as sharply, including ULSAM (-4%) and CHUP (-4%), which contributed to the decrease in the number of reference units and the overall average values of the CIs.

Conversely, CHPVVC stood out as the sole unit to exhibit a remarkable CI improvement from zero to one, signifying a 100% increase between 2019 and 2020, and maintaining this value in 2021. Analyzing the monthly figures, this improvement became evident starting in December 2019, marked by a significant improvement in the percentage of hospitalizations with a delay of more than 30 days, and in the percentage of patients operated within the TMRG. Moreover, ULSN also joined the roster of reference hospitals in 2020, with a 6.8% increase compared to the prior year, however it ceased to be one in 2021, due to an 8.6% decrease in its index value.

Based on the observed data, we can address the first research question (R1), concluding that Hospital Santa Maria Maior (HSMM) holds the highest quality index, followed by Hospital Distrital da Figueira da Foz (HDFF), Centro Hospitalar Entre Douro e Vouga (CHEDV), Unidade Local de Saúde de Alto Minho (ULSAM), and Centro Hospitalar do Médio Ave (CHMA). These hospitals align with the previously mentioned list of institutions that garnered the most benchmark classifications, with the exception of Centro Hospitalar Póvoa de Varzim/Vila do Conde (CHPVVC). Despite achieving reference status in 52.8% of the analyzed months, CHPVVC possesses an average quality index value of 0.67. Furthermore, Centro Hospitalar Universitário do Porto (CHUP) and Centro Hospitalar Tâmega e Sousa (CHTS) occupy the 6th and 7th positions in the quality ranking, both maintaining CI values above 0.9. Conversely, the units with the worst classification in the ranking are Unidade Local de Saúde do Litoral Alentejano (ULSLA), Centro Hospitalar Trás-os-Montes e Alto Douro (CHTMAD), and Centro Hospitalar Universitário Lisboa Norte (CHULN).

The following graphs depict the averages of the CIs throughout the study period. The graph on the right has been expanded to provide a clearer view of the range of difference values.

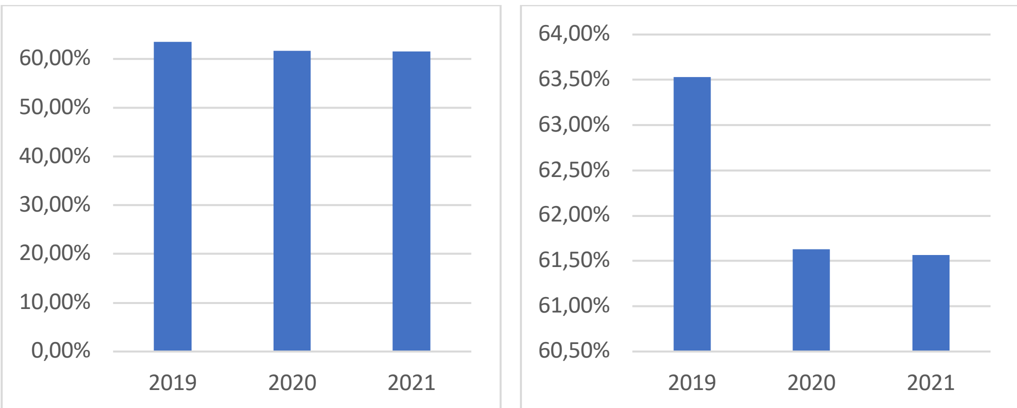

Figure 8- Average CIs over the analysis period (Source: Author).

There is a slight decrease in the CI values over the years. However, this decline is not statistically significant. In 2019, the average CI stood at 63.53%, followed by 61.63% in 2020 (a 3% decrease), and a further decrease to 61.57% in 2021 (a marginal 0.09% decrease). Additionally, in both 2020 and 2021, the number of entities with CI values between 0.9 and 1 declined, with a notable decrease occurring in 2020.

There is no way to discuss the year 2020 without acknowledging the profound impacts of COVID-19 pandemic and the measures taken to combat it. In Portugal, the effects of the pandemic became increasingly apparent at the outset of 2020, with the most challenging period extending from September 2020 to March 2021, when there was a peak in the number of hospitalizations and deaths caused by this virus. The highest peak in cases occurred in January 2021, largely attributed to the Christmas and New Year celebrations, as highlighted in a study by Caldas and Varela (2023). The fluctuation in COVID-19 hospitalizations during this critical period is illustrated in the graph below.

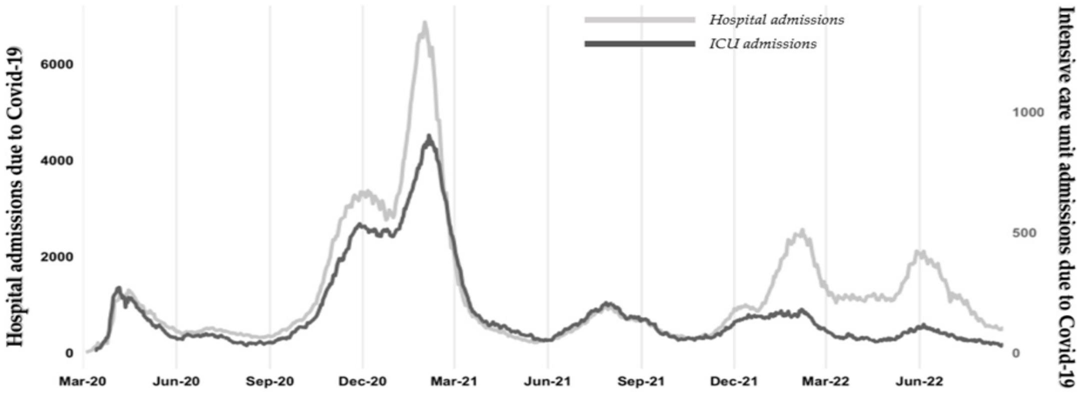

Figure 9- Hospital and intensive care units' admissions due to COVID-19, in Portugal (Source: DGS, Status Report no. 763[13]).

This period resulted in an overload of the health system, leading to a possible reduction in service capacity, which led to the redirection of resources to care for infected patients. Furthermore, the cancellation of non-urgent procedures and the emotional and psychological pressure on patients and healthcare professionals were notable, caused in part by the obligation to comply with visiting restrictions and social isolation. Thus, the dynamic nature of the pandemic forced rapid adaptation on the part of all healthcare entities, creating training and consistency challenges in the provision of care, potentially affecting its quality (Varanda et al., 2020).

To distinguish between hospitals with better performance (CI ≥ 0.9) and those with worse performance (CI < 0.9), an analysis of the average values of the KPIs for each of these two groups was conducted. It was observed that in 2019 the differences in results between the 2 groups were not very significant, making it challenging to discern clear patterns from the graph. This outcome was expected given the higher number of entities with good performance, leading to more similar means between the two groups. However, in 2020, the differences in the average values of the KPIs became more pronounced, and this trend continued into 2021. Therefore, for a more detailed analysis, we will focus on the 2021 values, as presented in Figure 10:

[13] DGS, Status Report no. 763. https://covid19.min-saude.pt/relatorio-de-situacao/, accessed on August 27, 2023.

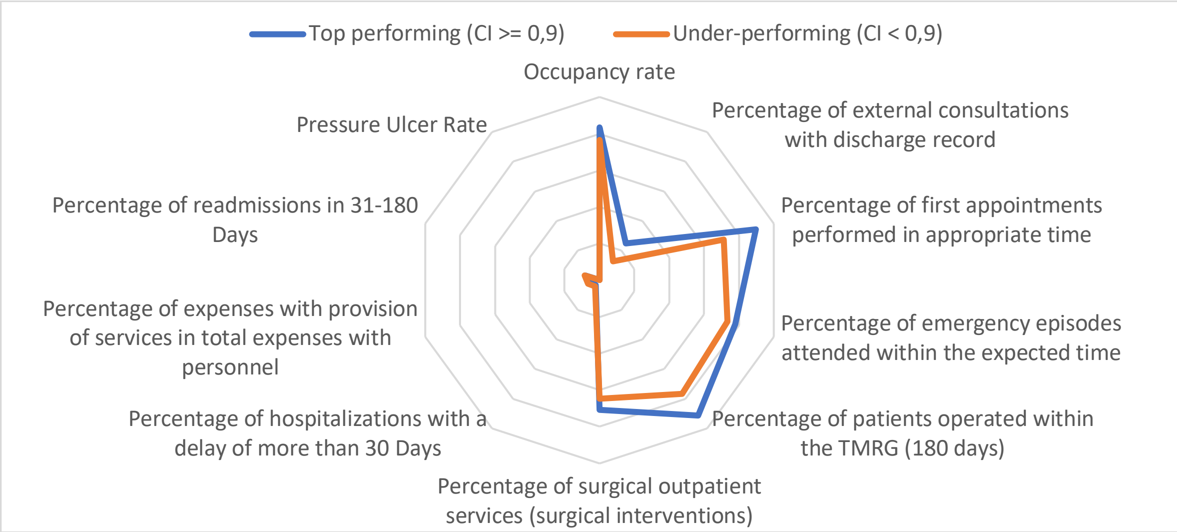

Figure 10- Comparison between average KPI values of top-performing and under-performing units (Source: Author).

This analysis clearly highlights that the group of top-performing hospitals has significantly higher average values for desirable indicators compared to the other group. However, concerning undesirable indicators, the average values are quite similar for both groups, suggesting that the hospitals' major challenges are not predominantly related to these indicators. Digging deeper into the desirable indicators, it becomes evident that those that stand out in the difference between the groups are the “Percentage of first appointments performed in appropriate time” and “Percentage of patients operated within the TMRG”, with a difference of 18.47 and 14.49 percentage points, respectively, indicating that top-performing hospitals perform notably better in these areas. Furthermore, despite being categorized as a desirable indicator, the indicator "Percentage of external consultations with discharge record" presents a significantly lower value than desired, representing an average of 24.67% for the group of top-performing hospitals and 12.77% for the other, which is almost half the value. This suggests that this particular indicator is a challenge for both groups, but it is even more pronounced for low-performing hospitals. In this context, healthcare entities aiming to enhance the quality of their services, particularly those ranking lower, should prioritize efforts directed at improving specific key areas, such as the percentage of first appointments performed in appropriate time, and the percentage of external consultations with discharge record. Although not prominently high, attention should also be directed towards reducing the percentage of readmissions occurring between 31-180 days, as this indicator surpasses the values of all other undesirable indicators.

To determine which category (hospital, HC, or ULS) exhibits superior performance, two distinct approaches were employed. Firstly, the average CI values were computed for entities within each category, in order to evaluate their performance relative to the others. Additionally, the proportion of entities achieving a general CI value exceeding 0.9, considering the number of entities in each group, was calculated, to verify the values obtained in the first approach.

Table 4- Performance comparison between health unit categories (Source: Author).

	CH	Hospital	ULS
Number of entities	21	8	8
Average CI	0,60	0,69	0,62
Proportion of reference entities	19%	25%	13%

Based on the information presented in the table, it is evident that the Hospital category exhibits the highest performance levels, followed by CHs, with ULS performing at a comparatively lower level.

5.3 Political and managerial implications of results

In the case of Hospital Santa Maria Maior (HSMM), the consistent achievement of benchmark-level results for three consecutive years prompts an investigation into the factors contributing to this exceptional performance. According to information available on the NHS website, HSMM embarked on an Accreditation process in December 2016, adhering to the model endorsed by the Ministry of Health in four hospital services: the Operating Block, Outpatient Surgery Unit, Clinical Pathology Laboratory, and Immunohemotherapy Service. This endeavor culminated in accreditation being conferred upon HSMM in July 2018, showing its commitment to prioritize quality, as evidenced by the satisfaction of its users, healthcare professionals, and other stakeholders. These remarkable results can be further underscored by the accolades received by HSMM. In 2020, it was distinguished as the best hospital in Group B of NHS hospitals (TOP 5) for its exceptional care and management performance in the year 2019. Notably, HSMM had previously earned the TOP5 designation in 2017, 2015, and 2014. The "TOP 5'20" ranking is administered by IASIST, a multinational organization of Spanish origin. This ranking is based on hospital benchmarking data from the Central Administration of the Health System and recognizes hospitals within the NHS that have consistently demonstrated the highest levels of performance throughout the year.

According to the approved action plan (disclosed as stipulated in no. 5 of Order no. 3635/2013, of March 7, from the Deputy Secretary of State to the Minister of Health), the vast majority of complaints presented to HSMM were associated with the waiting time in the Emergency Service, and the monitoring of the user in the Emergency Department (ED). Thus, in order to reduce waiting times in the ED, HSMM became involved in the NHS + Proximity project. Within this initiative, HSMM has taken part in a pilot program aimed at directing non-urgent patients to HCs, thereby alleviating the ED's workload. Furthermore, specific protocols have been implemented to regulate visitation and companionship within the emergency service. Notably, HSMM has achieved impressive results in indicators such as the "Percentage of emergency episodes attended within the expected time" and the "Percentage of first appointments performed in appropriate time." These results are well-aligned with the data used in this study, consistently showing high values in these variables throughout the analyzed period. Additionally, in 2021, HSMM surpasses all other entities with an annual average exceeding their performance in the "Percentage of first appointments performed in appropriate time." Moreover, HSMM significantly outperforms other units in terms of the "Percentage of external consultations with discharge records," with a substantial margin of 53.08 percentage points above the average for 2021. These achievements underscore the effectiveness of the measures implemented by HSMM in enhancing the quality of healthcare services.

Focusing now on the hospitals that present the worst value for the quality index, as well as the lowest frequency of being designated as reference hospitals, we find Centro Hospitalar Trás-os-Montes e Alto Douro (CHTMAD) and Unidade Local de Saúde do Litoral Alentejano (ULSLA) at the forefront of underperformance. Regarding the first hospital, a closer examination reveals that its subpar performance can be attributed to two key indicators: the "Percentage of first consultations carried out in adequate time" and the "Percentage of readmissions in 31-180 days." For the latter indicator, CHTMAD consistently records the lowest values across all years, indicating significant shortcomings in patient treatment and follow-up care. ULSLA, on the other hand, displays its weakest performance in the indicators "Percentage of first consultations carried out in adequate time", and "Percentage of patients operated within the scope of the TMRG." Particularly concerning is the second indicator, which exhibits the second-worst performance in 2019 and the worst performance in 2020. Additionally, the average values of the “Percentage of service provision expenses in total personnel expenses” are notably higher for ULSLA, compared to other health entities throughout the entire period.

It is important to acknowledge the challenges faced by CHTMAD, which integrates hospital units in Vila Real, Chaves, Lamego, and a Palliative Care unit in Vila Pouca de Aguiar. The region served by this hospital center is characterized by having one of the country's oldest populations, a geographically dispersed distribution, lower per capita income, and a tendency towards lower health literacy levels [14]. These contextual factors may contribute to the challenges faced by the healthcare institution in delivering quality services.

In turn, the 2020 ULSLA financial report, sourced from the NHS website, reveals critical challenges faced by this unit that negatively impacted its operational and financial performance. One of the significant challenges was related to patient waiting times for specific medical specialties, such as Orthopedics, Gynecology, and Urology, where patients had to wait more than 105 days to be intervened, in contrast to the 180 days foreseen according to the TMRG. This delay was primarily due to the high patient load in certain specialties, including Ophthalmology and Otorhinolaryngology. It is crucial to underscore that the overall percentages of patients receiving treatment within the TMRG are unsatisfactory for both types of medical conditions, with the oncological cases exhibiting worst values. The suspension of certain medical specialties for nearly a year due to a shortage of medical personnel resulted in a substantial number of patients remaining on the waiting list, even after their expected treatment times had lapsed.

The financial report also outlined challenges affecting ULSLA's goals. Firstly, the COVID-19 pandemic strained finances, impacting expenditures and revenues. It prompted the resource allocation from routine services to pandemic care, which hindered the ability to deliver regular healthcare services. USLA also struggles with maintaining a stable and adequately staffed workforce, aggravated by the substantial increase in absenteeism in 2020, resulting in an increase in overtime compared to the

[14] Portugal Clinical Trials, “Centro Hospitalar de Trás-os-Montes e Alto Douro”. https://portugalclinicaltrials.com/en/clinical-research-centers/centro-hospitalar-de-tras-os-montes-e-alto-douro/, accessed on September 20, 2023.

previous year. Furthermore, the institution faces a qualified staff shortage, compromising service quality, with some staff working on a limited basis, while others can only be supported through extensive overtime. In certain cases, specific services depend on just one doctor. The ULSLA administration recognized the urgent need to adapt the workforce to the demands of the local population in various areas, including doctors, nurses, health technicians, and administrative staff. To address this staffing crisis, ULSLA uses costly temporary providers, as budget limits hinder permanent staff expansion. The health facility's management is actively devising strategies to tackle these challenges, like attracting new talent, reshaping job roles, and seeking stable employment opportunities. This healthcare unit has, in fact, initiated measures to enhance healthcare services. Notably, an agreement with the Algarve Biomedical Center was formed in March 2022, for a five-year collaborative endeavor in the realms of research, training, and human resources. The primary objectives of this partnership include the enhancement of healthcare professional training, the advancement of clinical research initiatives, the provision of distinctive healthcare services, and the resource allocation for the development of innovative healthcare solutions.[15]

To enhance the performance of entities that have shown less satisfactory results, it is imperative to consider various mitigation measures aimed at improving efficiency, healthcare quality, and resource management. Therefore, it is important to mention not only the specific measures proposed by ULSLA, but also to outline strategies to promote better performance in healthcare provision, namely:

- Resource management optimization: ULSLA gives high priority to efficiency in resource allocation. This involves carrying out a comprehensive review of staffing, and leveraging technology to refine internal processes, identifying areas for operational improvement;
- Training and professional development: Investing in the training and continuous development of healthcare professionals. This includes training programs, skills upgrades, and the fostering of a continuous learning environment among healthcare professionals;
- Redistribution of tasks and services: Carry out meticulous reevaluation of the allocation of responsibilities and services among different teams. Ensure that roles are allocated judiciously, optimizing service capacity across the organization;
- Assessment of community health needs: An ongoing analysis of the community's health needs is essential. This includes a demand analysis to target resources according to the specific health needs of the population, assessment of health resources available in the community, analysis of epidemiological data to identify health trends in the community (such as prevalent diseases, at-risk demographic groups, and emerging risk factors), assessment of the quality of existing health services, as carried out in this work, among others;
- Strategic partnerships: ULSLA is already implementing this strategy, as previously mentioned, which involves establishing strategic partnerships with other health entities, academic institutions, and health-related organizations. These collaborations are aimed at

[15] Postal do Algarve, "ABC assina protocolo com Unidade Local de Saúde do Litoral Alentejano para investigação e melhoria dos cuidados de saúde". https://postal.pt/saude/abc-assina-protocolo-com-unidade-local-de-saude-do-litoral-alentejano-para-investigacao-e-melhoria-dos-cuidados-de-saude/, accessed on September 20, 2023

fortifying the organization's resource base and capabilities, facilitating a more robust response to healthcare challenges.

5.4 Impact of sociodemographic variables on hospital performance

To address the second research question in this study, an in-depth analysis of exogenous variables and their potential influence on the CIs outcomes was undertaken. This research began with a rigorous examination of multicollinearity among the independent variables. This step was crucial to ensure that these variables did not exhibit high correlations, which could lead to result distortions and hinder precise interpretations of each variable's individual contributions. To this end, the Variance Inflation Value (VIF) was used, a statistical metric that quantifies the extent to which regression coefficients' variances are inflated. A VIF value of one denotes an absence of multicollinearity, whereas values exceeding one indicate potential presence of this problem, noting that, typically, a VIF equal to or greater than ten is considered indicative of significant multicollinearity. During this analysis, certain variables, such as "Beds," "Population Density" and "Education," exhibited VIFs exceeding ten, which can be elucidated by the intrinsic relationship between the population size, the number of students, and hospital bed capacity, as more densely populated regions tend to necessitate more extensive healthcare services, thereby requiring greater hospital resources.

As such, one effective approach for dealing with multicollinearity is the construction of alternative models, testing different combinations of independent variables, in order to determine which variables are most relevant in explaining the dependent variable. Several techniques can be employed for variable selection, and in this study the stepwise backward regression method has been chosen. This method is well-known for its iterative process of adding and removing predictors within the predictive model, seeking to find the subset of variables that results in a better performing model, i.e., the one with the lowest precision error. In addition to this, it helps mitigate multicollinearity issues. This method was selected to perform the multiple linear regression analysis because, in comparison to the "stepwise" and "stepwise forward" methods, it resulted in a model with a greater number of independent variables that exhibited statistical significance, indicating that these variables played an essential role in explaining the outcomes of the CIs.

In 2019, when applying the method described previously, it was evident that none of the exogenous variables presented statistical significance, as they all presented a p-value greater than 0.05. This observation suggests that, during that year, exogenous variables did not exert a significant impact on the overall quality outcomes. The results of this analysis are summarized in table 5.

Table 5-Significance of variables in regression analysis for the CIs in 2019 (Source: Author).

Variable	Estimation (coefficient)	p-value	Std. Error	t value
(Intercept)	0,580	0,556	0,974	0,596
Beds	0,000	0,945	0,000	0,070
AgingIndex	0,002	0,525	0,003	0,645
PopulationDensity	0,000	0,606	0,000	-0,522
TeachingStatus	-0,377	0,156	0,258	-1,460
Education	-0,002	0,978	0,076	-0,028
Mortality	-0,011	0,800	0,044	-0,255
PatientsperDoctor	0,000	0,684	0,001	0,412
DoctorsPerInhabitants	0,049	0,140	0,032	1,520

However, when moving forward to 2020, the variable selection process identified a model with two variables that proved to be statistically significant: the number of hospital beds (Beds) with a p-value of 0.01 and the educational level (Education), with a p-value of 0.014. These results indicated that, in this year, these two variables had a statistically significant impact on the CIs. When examining the term "intercept," we understand that this effect is positive for the level of education and negative for the number of beds.

Table 6-Significance of variables in regression analysis for the CIs in 2020 (Source: Author).

Variable	Estimation (coefficient)	p-value	Std. Error	t value
(Intercept)	0,965	0,001	0,273	3,531
Education	0,097	0,014	0,037	2,605
Beds	-0,001	0,010	0,000	-2,750
PatientsperDoctor	0,000	0,697	0,001	-0,393
DoctorsPerInhabitants	-0,024	0,217	0,019	-1,259

Interpreting these findings, the positive correlation between the educational level and the CI values implies that higher educational attainment within the population is associated with improved quality outcomes in hospitals. This observation can be attributed to enhanced patient comprehension of healthcare processes, and an increased ability to adhere to treatments and medical recommendations. Conversely, the negative correlation between the number of hospital beds and CIs suggests that a higher bed count does not necessarily lead to superior quality results, what makes a certain sense, as hospitals with a larger number of beds often encounter more complex challenges, such as resource management, impacting directly on KPIs.

As the study progressed to 2021, four variables that held statistical significance were identified: the density of doctors per 1000 residents (Doctors per inhabitants), the teaching status of hospitals (Teaching status), the rate of mortality (Mortality), and lastly, the number of patients per doctor (Patients per doctor), which the program assumed a very small level of significance, even though it had a p-value greater than 0.05.

Table 7- Significance of variables in regression analysis for the CIs in 2021 (Source: Author).

Variable	Estimation (coefficient)	p-value	Std. Error	t value
(Intercept)	0,773	0,011	0,288	2,684
PatientsperDoctor	0,001	0,089	0,001	1,753
DoctorsPerInhabitants	0,036	0,050	0,018	2,040
TeachingStatus	-0,351	0,031	0,155	-2,259
Mortality	-0,045	0,016	0,018	-2,546

Analyzing the coefficient values in the table, it is observed that the first two variables have a positive influence on the values of the CIs, meaning that an increase in these variables can potentially lead to an increase in quality. Conversely, the last two variables exhibit negative coefficients, meaning that they have a negative impact on the dependent variable. The variable "Mortality" aligns with expectations, showcasing an inverse relationship with the CI. Thus, it tends to correspond with poorer performance in the CI, a logical connection given that elevated mortality rates are typically associated with healthcare issues. However, the direction of impact for the "TeachingStatus" variable may initially appear counterintuitive. While one might anticipate that teaching hospitals, with their focus on education and research, would yield better outcomes, a negative coefficient, in this context, suggests that teaching hospitals could face additional challenges that negatively affect the CIs, such as increased case complexity due to medical teaching and research.

The R^2 (R-squared) value in the context of multiple linear regression represents the proportion of variability in the dependent variable that is explained by the independent variables included in the model. In this instance, an R^2 of 0.232 was obtained, which means that the independent variables included in this model explain approximately 23.2% of the variability in the value of the CI. Although this value suggests that the model possesses some capacity to explain variations in the dependent variable, there is still a substantial portion of the variability that remains unexplained by the independent variables within the model. Consequently, there are other factors, not considered in this model, that can also exert influence on CI results.

The variation in the significance of the independent variables over the years, notably the increase in 2020 and 2021 compared to 2019, may be due to many factors. Primarily, the COVID-19 pandemic's impact in 2020 and more so in 2021 drastically affected the health sector and the variables influencing hospital KPIs. Increased mortality, pressure on health systems, and the need for rapid adaptations in health policies potentially impacted the "Mortality" variable on quality assessments. Secondly, population behavior and healthcare policy adjustments, in response to the pandemic, could explain changes in 2020 and 2021. The “Teaching status” variable might have gained added significance during this period, as hospitals often play pivotal roles in managing crises, typically having more resources, highly specialized medical staff, and serving as research centers. Therefore, with the advent of COVID-19, the importance of teaching hospitals may have surged, resulting in varied outcomes in the CIs. However, the negative coefficient suggests that teaching hospitals faced unique challenges or handled more complex cases. The pandemic might have also altered public health awareness and service access, influencing the significant relationship with doctor density per inhabitant (DoctorsPerHabitants), indicating that regions with more doctors had better pandemic responses and KPIs.

When analyzing the variables that demonstrated significance in 2020, the variables "Education" and "Beds" stood out. The variable "Education" showed a significant positive coefficient, suggesting that, in 2020, regions with a higher level of education in the population tended to have improved hospital KPIs. This outcome is logical, as individuals with greater educational attainment often possess better health literacy and are more knowledgeable about health-related matters. Consequently, these regions may have witnessed more effective public health communication, leading to increased adherence to preventive measures, such as social distancing and mask-wearing, what likely contributed to reduced virus transmission, thereby positively influencing hospital outcomes. On the other hand, the variable "Beds" demonstrated a significant negative relationship with the CIs in 2020, implying that, during that year, an increase in the number of beds corresponded to a decrease in quality outcomes. This finding can also be explained in light of the pandemic context, where the surge in hospital bed capacity was primarily driven by the urgent need to accommodate patients with COVID-19. In many cases, a high number of beds could be related to the need to accommodate infected patients, indicating a substantial challenge to the overall quality of care.

5.5 Sensitivity analysis

In the initial BoD model used to calculate the CIs, restrictions were defined on the weights of the indicators, placing a lower limit of 0.05, as dictated by Eq. (8). This stipulated that every indicator should have a minimum contribution of at least 5% to the CI computation, and an upper threshold of one. Given that the total sum of weights must always be equal to one, and there are ten indicators, the maximum contribution each can have is 0.10, thus the upper limit did not impact the results.

Therefore, to analyze the sensitivity of the model, the restriction was eliminated, considering the null value as the lower limit, in order to observe possible changes in the CI results. Consequently, it was observed that the vast majority of DMUs received null weights for several indicators, disregarding their contributions to the CI calculation, which led to very high results. Thus, considering the 2019 values, instead of 13 DMUs referenced as benchmarks (CI=1), there were now 30, and the remaining ones also presented very high CI values. This indicates that, without adequate restrictions, the model may attribute disproportionate importance to certain indicators to the detriment of others.

In subsequent analysis, the lower limit constraint was increased to values ranging between 0.06 and 0.10. In this attempt to impose a higher minimum contribution for each indicator, there is a tendency for CI values to decrease as the limit increases, although there are some exceptions. However, this change is not drastic, suggesting that the model is sensitive, but to some extent, to changes in weight constraints. The values obtained for both sensitivity analyzes are presented in table 15, in the Appendix.

6. Conclusions and future work

6.1 Conclusive summary

The NHS in Portugal, established in 1979, is a universal public healthcare system designed to provide medical services to citizens, regardless of their economic or social status. It provides a wide range of health services, and operates as a tax-funded healthcare system, resembling a Beveridge model. However, this system grapples with challenges related to efficient resource management, due to its complexity and political changes, with the most pressing problem not solely being inadequate funding, but also concerns about how existing resources are deployed. In recent years, there has been an escalation in the demand for healthcare services, driven by a lack of resources and an aging population, which increases the pressure on Portuguese hospitals to maintain high-quality care despite financial limitations. In 2023, the weight of the health sector in the state's primary expenditure is amplified to 13.9%. Consequently, policymakers and healthcare administrators face the urgent task of finding innovative solutions to manage this increasingly unsustainable expenditure and address the lack of critical data. Thus, search for quality has become a central objective in most health systems, as hospitals are under great pressure to excel in their operations, particularly in resource utilization, and in the quantity and quality of outcomes achieved.

Hence, this study aims to contribute to the evaluation of healthcare quality and services in Portuguese hospitals. The assessment of care and treatment processes' quality plays a pivotal role in research, clinical practice, and healthcare evaluation. This represents the initial step in fostering quality and enhancing organizational efficiency (Gomes et al., 2020). There is a huge number of studies in the literature dedicated to measure or evaluate the performance of healthcare organizations, and a predominant approach in this domain is the use of DEA. Since this methodology allows the construction of models with many inputs and outputs, it becomes the most attractive for the study of this type of organizations, as they have multiple resources, facilitating this process. Moreover, DEA stands out for its capacity to utilize all available data to establish what is known as the BPF, which represents the level of maximum performance achievable by a hospital, given its inputs and outputs, enabling contextualized projections to address specific challenges in the healthcare domain.

In this study, a distinct approach to the DEA model, known as BoD, was employed. Unlike certain other performance evaluation techniques that primarily focus on the transformation of resources into products, BoD adopts a different perspective. Its objective is to aggregate multiple individual KPIs into a single performance measure, without explicitly emphasizing inputs. Thus, there are several compelling reasons for choosing this method over others, particularly for the creation of CIs. Firstly, it enables direct comparisons with all other observations within the sample, making it highly valuable for evaluating the relative performance of various units of analysis. Secondly, the method permits the specification of multipliers through an optimization process, leading to a more precise allocation of weights to performance indicators. Lastly, BoD offers a high degree of flexibility in accommodating variables with different units of measurement, eliminating the necessity for prior data normalization, which facilitates the inclusion of a variety of indicators, even if they measure different aspects of performance.

One of the determining components for this study involved the meticulous selection of variables to be used as performance indicators. To accomplish this, an extensive literature review was conducted, wherein the most frequently employed KPIs, as per existing research, were compiled. Consequently, a set of ten indicators was identified for gauging the quality of healthcare services, consisting of six desirable and four undesirable variables. To accommodate the varying nature of these indicators, the min-max method was employed, which is a normalization technique that allows the assignment of both positive and negative polarities to the indicators. It essentially standardizes all the indicators onto a uniform scale, wherein zero represents the poorest possible performance and one symbolizes the utmost level of performance attainable, simplifying the process of comparing and aggregating the indicators into a unified CI. Therefore, the analysis of the quality of Portuguese public hospitals was conducted for the years 2019, 2020, and 2021.

Within the realm of healthcare research, it is notable that a considerable majority of studies have predominantly centered around efficiency analysis, with comparatively fewer addressing quality evaluation. Consequently, this study assumes a distinctive position owing to its methodology and its incorporation of quality assessment indicators, specific to public healthcare institutions in Portugal.

With these considerations in mind, the comprehensive analysis conducted in this study has yielded several key findings and conclusions. These conclusions are offered in a condensed form for brevity, and readers seeking a more comprehensive understanding are encouraged to refer to the preceding chapter for an in-depth exploration of the findings.

- Regarding the type of organization, hospital entities of the “Hospital” type stood out the most in terms of performance, followed by HCs and ULS.
- The hospital entity with the best performance was Hospital Santa Maria Maior (HSMM), emerging as the sole benchmark entity across all three years under examination.
- The entity most often considered as a reference, according to the monthly analysis of the CIs, was Centro Hospitalar Póvoa de Varzim/Vila do Conde (CHPVVC), presenting a benchmark value in 52.8% of the months, over the three-year span.
- In second place in the ranking, Hospital Distrital da Figueira da Foz (HDFF) must be distinguished, with a minimal difference compared to top-performing entity and achieving a reference status 44.4% of the time.
- Regarding the variation in quality throughout the period analyzed, there was a decrease in both the number of reference entities and the average CI values. However, the latter decrease was not particularly significant.
- Regarding the analysis of the influence of sociodemographic variables on the quality of hospitals, in the year 2019, the analysis did not reveal any significant impact of these variables on the CI values. However, in 2020, two variables – Beds and Education – were found to have a significant influence on the healthcare quality. Finally, in 2021, the program identified four variables – Doctors per Inhabitants, Patients per doctor, Teaching Status and Mortality – as factors that had a discernible impact on the dependent variable (CI).

6.2 Limitations

Upon the completion of this research work, it is imperative to acknowledge its inherent limitations. These limitations are crucial to consider when scrutinizing the obtained results, as they can significantly impact the accuracy and scope of the interpretations drawn from this study.

First and foremost, it is crucial to acknowledge that data availability posed a significant challenge in this study. The initial intention was to conduct a performance analysis that included the most recent data up to the year 2022. However, due to data limitations in certain variables, this objective could not be fully realized. Furthermore, consistent access to comprehensive and up-to-date data was not achieved across all variables of interest present in the study, which may have limited our analyzes and restricted our ability to draw extensive conclusions in specific areas of the study. In some cases, missing or inconsistent data were identified in the data sources used, which may have affected the integrity and reliability of the analyzes conducted. It is also worth noting that some variables presented values greater than 100%, resulting in a possible overestimation of positive performance, or in the case of undesirable variables, the opposite. In cases where this occurred, values exceeding 100% were replaced by the respective indicator's average, which unfortunately led to the loss of information and, subsequently, less precise conclusions.

Regarding the sample used in this study, it must be recognized that it may not offer a fully representative image of the target population, primarily a result of restrictions related to data accessibility and specific selection criteria. In particular, the exclusion of certain entities specializing in healthcare, such as Portuguese oncology institutes, regional alcoholology centers, orthopedic and ophthalmological facilities, carries the inherent risk of restricting the generalizability of our results and introducing potential biases in the assessed performance levels.

Multicollinearity among some of the exogenous variables emerged as a noteworthy concern. This high degree of correlation between variables presents challenges in isolating their individual contributions to the study outcomes. Therefore, to mitigate this issue, the variables were segmented into distinct models. However, this approach could potentially introduce biases into the results. Furthermore, concerning these variables, despite extensive research utilizing sources such as PORDATA and INE, we encountered instances of missing or null values for several municipalities. This lack of data completeness reduced the statistical significance of these variables, making it difficult to discern their influence on the CI values. Finally, still regarding the search for sociodemographic variables, most of the available data is aggregated at the national level, with a lack of specific information at the municipal level. This scarcity hampered the ability to incorporate important variables into the model, which were initially considered vital to strengthen the depth of the analysis.

In addition to the limitations arising from data restrictions, it is essential to highlight those associated with the model itself. The CI values obtained depend on (a) the specific sample under consideration, (b) the selection of variables used as KPIs, and (c) the restrictions imposed on the multipliers (Greco et al., 2019). Thus, it should be noted that any changes in these aspects have the potential to generate results that differ significantly from those presented in this study. Furthermore, the BoD methodology operates under the assumption of compensability between indicators, allowing low values in certain indicators to be compensated by those with higher values. This characteristic was evident by the existence of numerous zeros in the initial set of multipliers obtained.

Finally, the methodology used to analyze performance evaluates it in a relative way, considering the DMUs investigated. As a result, even entities classified as benchmarks, when compared with others in the analysis, must remain attentive in encouraging the adoption of practices that aim to improve their performance. Being considered top-performing entities in a sample analysis does not mean that they have reached the peak of their potential and have no room for improvement in different contexts.

6.3 Future work

The intention of this work was to enrich benchmarking studies, providing innovative, comprehensive, and more complete research, with a special focus on the hospital sector. However, the final results presented are not definitive, so it is recommended to compare these results with subsequent studies that use more recent data, and potentially explore new performance groups. Therefore, to strengthen the robustness of the results obtained, it is worth highlighting some recommendations for future research.

Firstly, encourage the implementation of additional studies in the healthcare field, particularly within hospital entities, that include a variety of quality indicators, beyond those conventionally associated with efficiency. Moreover, it is also important to incorporate new indicators that cover different categories, enabling a more comprehensive analysis and facilitating broader comparisons. Additionally, the inclusion of excluded entities in this work, due to imperfect knowledge of the data, would also be interesting to validate the results presented.

Based on the analysis performed in this dissertation, a noteworthy recommendation for future research is the implementation of a cluster analysis at the beginning of the study. This approach would facilitate the creation of homogeneous groups by including hospitals that share high similarity based on the variables used. Subsequently, intercluster comparisons could be carried out, for example using the k-means method.

Furthermore, it would be pertinent to conduct temporal trend analyses for the indicators under scrutiny. Approaches like the Malmquist index could be employed to gain a more profound understanding of how the performance of the entities in question has evolved over the years.

A recent article in "Jornal de Negócios" highlights a notable shift in Portugal, where an increasing number of citizens are opting for private health services as their primary choice .[16] This trend is not solely attributed to the accessibility and efficiency of private healthcare, but also reflects a growing confidence in its quality. Meanwhile, the public health system has encountered significant challenges, including a gradual erosion of its position to the private sector, which often recruits medical professionals from the public sector, taking advantage of the precarious working conditions, and lack of motivation in this environment. Given this scenario, Portuguese citizens aspire to have access to timely, high-quality healthcare, a goal that can only be achieved through synergistic and complementary collaboration

[16] Negócios. (2023). “Público vs. privado? 'Todos temos um papel a desempenhar'.” Portugal Health Summit. https://www.jornaldenegocios.pt/negocios-iniciativas/portugal-health-summit/detalhe/publico-vs-privado-todos-temos-um-papel-a-desempenhar, accessed on September 23, 2023.

between the public, private, and social sectors within healthcare. In this sense, a highly relevant analysis would be to understand the factors contributing to this trend, including the role of public-private partnerships, in order to identify the specific areas where the public health sector can enhance its services, based on good practices observed in the private sector.

Furthermore, expanding the scope of this analysis to include hospitals from other countries is a crucial consideration, which would provide an international context, offering a more comprehensive and insightful perspective on areas requiring improvement and successful strategies that have been implemented globally.

Although this study carried out an analysis to evaluate the influence of exogenous factors on healthcare entities' performance, future research should explore various analytical tools, including m-order models. These approaches can further enrich our understanding of the intricate dynamics within healthcare systems and open up avenues for more nuanced insights.

Lastly, given the absence of university hospitals identified as benchmarks in the overall performance analysis, and the significant negative coefficient associated with the "teaching status" variable in 2021, suggesting a potential adverse impact of teaching activities on hospital performance, further in-depth investigation is warranted. Specifically, exploring the possible relationship between the value achieved for the aggregate quality indicator and the involvement of hospital institutions in teaching activities would be a valuable avenue for future research.

References

Afonso, A., & Fernandes, S. (2008). Assessing Hospital Efficiency: Non-Parametric Evidence for Portugal. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1092135

Aloh, H. E., Onwujekwe, O. E., Aloh, O. G., & Nweke, C. J. (2020b). Is bed turnover rate a good metric for hospital scale efficiency? A measure of resource utilization rate for hospitals in Southeast Nigeria. *Cost Effectiveness and Resource Allocation*, 18(1).

Amado, C. A., & Santos, S. P. (2009). Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres. *Health Policy*, 91(1), 43–56.

Arocena, P., & García-Prado, A. (2007). Accounting for quality in the measurement of hospital performance: evidence from Costa Rica*. Health Economic*s, 16(7), 667–685.

Ashton, C. M., Kuykendall, D. H., Johnson, M. L., Wray, N. P., & Wu, L. (1995). The association between the quality of inpatient care and early readmission. *Annals of Internal Medicine*, 122(6), 415-421.

Backman, C., Vanderloo, S., & Forster, A. J. (2016). Measuring and improving quality in university hospitals in Canada: The Collaborative for Excellence in Healthcare Quality. *Health Policy*, 120(9), 982–986.

Balla, U., Malnick, S., & Schattner, A. (2008). Early readmissions to the department of medicine as a screening tool for monitoring quality of care problems. *Medicine,* 87(5), 294-300.

Barbetta, G. P., Turati, G., & Zago, A. M. (2006). Behavioral differences between public and private not-for-profit hospitals in the Italian national health service. *Health Economics*, 16(1), 75–96.

Barros, C. P., de Menezes, A. G., & Vieira, J. C. (2011). Measurement of hospital efficiency, using a latent class stochastic frontier model. *Applied Economics*, 45(1), 47–54.

Barros, C. P., Menezes, A.G., Peypoch, N., Solonandrasana, B., & Vieira, J. C. (2007). An analysis of hospital efficiency and productivity growth using the Luenberger indicator. *Health Care Management Science*, 11(4), 373–381.

Blank, J. L. T., & Valdmanis, V. G. (2009). Environmental factors and productivity on Dutch hospitals: a semi-parametric approach. *Health Care Management Science*, 13(1), 27–34.

Bogetoft, P., & Otto, L. (2011). *Benchmarking with DEA, SFA, and R. In International Series in Operations Research & Management Science.* Springer New York. https://doi.org/10.1007/978-1-4419-7961-2

Borden, J.P. (1988). An assessment of the impact of diagnosis-related group (DRG)-based reimbursement on the technical efficiency of New Jersey hospitals using data envelopment analysis. *Journal of Accounting and Public Policy*, 7(2), 77–96.

Braillon, A., Chaine, F., & Gignon, M. (2008). Le Benchmarking, une histoire exemplaire pour la qualité des soins [Benchmarking, an exemplary story for the quality of care]. *Annales Françaises d'Anesthésie et de Réanimation,* 27(6), 467–469.

Bull, A. (1994). Specifying quality in health care. *Journal of Management in Medicine*, 8(2), 5-8.

Burgess, J., & Wilson, P.W. (1995). Decomposing hospital productivity changes, 1985-1988: A non-parametric Malmquist approach. *Journal of Productivity Analysis*, 6(4), 343–363.

Burstin, H.R., Conn, A., Setnik, G., Rucker, D.W., Cleary, P.D., O'Neil, A.C., Orav, E.J., Sox, C.M., & Brennan, T.A. (1999). Benchmarking and quality improvement: The Harvard emergency department quality study. *American Journal of Medicine*, 107(5), 437–449.

Calabria, F. A., Camanho, A. S., & Zanella, A. (2016). The use of composite indicators to evaluate the performance of Brazilian hydropower plants. *International Transactions in Operational Research*, 25(4), 1323–1343.

Caldas, P., & Varela, M. (2023). The Portuguese Public Hospitals Performance Evolution before and during the SARS-CoV-2 Pandemic (2017–2022). *Sustainability*, 15(15), 11572.

Camp, R.C., & Tweet, A.G. (1994). Benchmarking applied to health care. *The Joint Commission Journal on Quality Improvement*, 20(5), 229-238.

Campbell, S.M., Braspenning, J., Hutchinson, A., & Marshall, M.N. (2003). Improving the quality of health care: Research methods used in developing and applying quality indicators in primary care. *BMJ*, 326(7393), 816–819.

Canilho, J. I. da F. (2019). *Análise da eficiência dos Hospitais Públicos Portugueses com Recurso à metodologia de Data Envelopment Analysis.* [Unpublished master's dissertation]. Universidade de Lisboa.

Carey, K. (2003). Hospital Cost Efficiency and System Membership. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 40(1), 25–38.

Carrilo, M., & Jorge, J. M. (2017). DEA-like efficiency ranking of regional health systems in Spain. *Social Indicators Research,* 133, 1133–1149.

Castro, R.A. d. S. (2011). *Benchmarking de Hospitais Portugueses - Modelação com Data Envelopment Analysis.* [Unpublished master's dissertation]. Universidade do Porto.

Castro, R. A. S., Portela, C.S., & Camanho, A. S. (2014). *Benchmarking dos serviços dos hospitais portugueses: uma aplicação de data envelopment analysis. In R. C. Oliveira & J. S. Ferreira (Eds.), Investigação operacional em ação: casos de aplicação* (pp. 703-740) [Operational investigation in action: application cases]. Imprensa da Universidade de Coimbra.

Charnes, A., Cooper, W.W., Lewin, A. Y., & Seiford, L. M. (1995). *Data Envelopment Analysis: Theory, Methodology, and Applications*. Springer Dordrecht. https://doi.org/10.1007/978-94-011-0637-5

Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), 429–444.

Chang, S., Hsiao, H. C., Huang, L. H., & Chang, H. (2011). Taiwan Quality Indicator Project and hospital productivity growth. *Omega*, 39(1), 14–22.

Chenoweth, J. (2003). *Benchmarking could save hospitals billions: Mortality, complications could be reduced.* Relias Media. Consulted on September 13, 2023. https://www.reliasmedia.com/articles/27063-benchmarking-could-save-hospitals-billions

Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T.V. (2007). An Introduction to "Benefit of the Doubt" Composite Indicators. *Social Indicators Research*, 82(1), 111–145.

Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T.V. (2011). Constructing composite indicators with imprecise data: A proposal. *Expert Systems with Applications,* 38(9), 10940–10949.

Chowdhury, H., & Zelenyuk, V. (2016). Performance of hospital services in Ontario: DEA with truncated regression approach. *Omega*, 63, 111–122.

Cole, M. (2009). Benchmarking: A process for learning or simply raising the bar? *Evaluation Journal of Australasia*, 9(2), 7-15.

Collins-Fulea, C., Mohr, J.J., & Tillett, J. (2005). Improving midwifery practice: The American College of Nurse-Midwives' benchmarking project. *Journal of Midwifery & Women's Health*, 50(6), 461–471.

Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA) – Thirty years on. *European Journal of Operational Research*, 192(1), 1–17.

Cooper, W.W., Seiford, L.M., & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References, and DEA-Solver Software*. Springer US. https://doi.org/10.1007/978-0-387-45283-8

Costa, B. B. (2021). *Fatores que determinam a eficiência produtiva dos sistemas de saúde de alguns países da OCDE.* [Unpublished master's dissertation]. Universidade do Porto.

Costa Jr, A. S., Leão, L. E. V., Novais, M. A. P., & Zucchi, P. (2015). An assessment of the quality indicators of operative and non-operative times in a public university hospital. *Einstein (São Paulo),* 13(4), 594–599.

Dismuke, C.E., & Sena, V. (1999). Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portuguese public hospitals? An empirical analysis using parametric and non-parametric methods. *Health Care Management Science*, 2(2), 107-116.

Donabedian, A. (1980). *Explorations in quality assessment and monitoring*, *Volume 1: The definition of quality and approaches to its assessment.* Health Administration Press. https://philpapers.org/rec/DONEIQ

Donabedian, A. (1988). The quality of care. How can it be assessed? *JAMA: The Journal of the American Medical Association*, 260(12), 1743–1748.

Donabedian, A. (2005). Evaluating the Quality of Medical Care. *Milbank Quarterly*, 83(4), 691–729.

Dyson, R., Allen, R., Camanho, A., Podinovski, V., Sarrico, C., & Shale, E. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2), 245–259.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), 253–290.

Ferrando, A., Ivaldi, C., Buttiglieri, A., Pagano, E., Bonetto, C., Arione, R., Scaglione, L., Gelormino, E., Merletti, F., & Ciccone, G. (2005). Guidelines for preoperative assessment: Impact on clinical practice and costs. *International Journal for Quality in Health Care*, 17(4), 323–329.

Ferrari, A. (2006). Market-oriented reforms of health services: A non-parametric analysis. *Services Industries Journal*, 26(1), 1–13.

Ferreira, D. C., & Marques, R. C. (2019). Do quality and access to hospital services impact on their technical efficiency? *Omega*, 86, 218–236.

Ferreira, D.C., & Marques, R.C. (2021). Public-private partnerships in health care services: Do they outperform public hospitals regarding quality and access? Evidence from Portugal. *Social Science & Economic Review*, 73, 100798.

Ferreira, D.C., Figueira, J.R., Greco, S., & Marques, R.C. (2023). Data Envelopment Analysis models with imperfect knowledge of input and output values: An application to Portuguese public hospitals. *Expert Systems with Applications*, 231, 120543.

Ferreira, D.C., Marques, R.C., & Nicola, P. (2013). On evaluating health centers groups in Lisbon and Tagus Valley: Efficiency, equity, and quality. *BMC Health Services Research*, 13(1), 529.

Ferreira, D. C., Nunes, A. M., & Marques, R. C. (2018a). Doctors, nurses, and the optimal scale size in the Portuguese public hospitals. *Health Policy*, 122(10), 1093–1100.

Ferreira, D.C., Marques, R.C., & Nunes, A.M. (2018b). Economies of scope in the health sector: The case of Portuguese hospitals. *European Journal of Operational Research*, 266(2), 716–735.

Ferreira, D. C., Marques, R. C., Nunes, A. M., & Figueira, J. R. (2018c). Patients' satisfaction: The medical appointments valence in Portuguese public hospitals. *Omega*, 80, 58–76.

Ferreira, D. C., Nunes, A. M., & Marques, R. C. (2019). Optimizing payments based on efficiency, quality, complexity, and heterogeneity: the case of hospital funding. *International Transactions in Operational Research*, 27(4), 1930–1961.

Ferreira, D. C., Nunes, A. M., & Marques, R. C. (2020). Operational efficiency vs clinical safety, care appropriateness, timeliness, and access to health care. *Journal of Productivity Analysis*, 53(3), 355–375.

Ferrier, G. D., & Trivitt, J. S. (2012). Incorporating quality into the measurement of hospital efficiency: a double DEA approach. *Journal of Productivity Analysis*, 40(3), 337–355.

Flokou, A., Aletras, V., & Niakas, D. (2017). A window-DEA based efficiency evaluation of the public hospital sector in Greece during the 5-year economic crisis. *PLoS ONE*, 12(5), 1–26.

French, S. (1995). Uncertainty and Imprecision: Modelling and Analysis. *The Journal of the Operational Research Society*, 46(1), 70–79.

Gaaloul, H., & Khalfallah, S. (2014). Application of the "Benefit-of-the-Doubt" approach for the construction of a digital access indicator: A revaluation of the "digital access index". *Social Indicators Research,* 118(1), 45-56.

Garcia-Lacalle, J., & Martin, E. (2010). Rural vs urban hospital performance in a 'competitive' public health service. *Social Science & Medicine*, 71(6), 1131-1140.

Gibari, S. El, Gómez, T., & Ruiz, F. (2019). Building composite indicators using multicriteria methods: a review. *Journal of Business Economy*, 89, 1–24.

Giraldes, M. do R. (2007). Avaliação da eficiência e da qualidade em hospitais: Entidades Públicas Empresariais e Sector Público Administrativo. *Acta Médica Portuguesa*, 20, 471-490.

Golany, B., & Roll, Y. (1989). An application procedure for DEA. *Omega*, 17(3), 237-250.

Gomes, E. P. do V. P. (2015). *Eficiência das unidades hospitalares no âmbito dos novos modelos de gestão: aplicação da metodologia Data Envelopment Analysis*. [Unpublished master's dissertation]. Instituto Superior de Contabilidade e Administração de Aveiro

Gomes, J. A. P., Martins, M. M., Tronchin, D., & Fernandes, C. S. (2020). Operating Room Quality in Portuguese Hospitals. *Perioperative Care and Operating Room Management*, 18, 100086. https://doi.org/10.1016/j.pcorm.2020.100086.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141, 61–94.

Groocock, J., Wiley, J. W., & Baker, A. (1988). The Chain of Quality. *R & D Management*, 18(3), 291–291.

Grosskopf, S., Margaritis, D., & Valdmanis, V.G. (2004). Competitive effects on teaching hospitals. *European Journal of Operational Research*, 154(2), 515–525.

Grosskopf, S., & Valdmanis, V. (1987). Measuring hospital performance. *Journal of Health Economics*, 6(2), 89–107.

Halfon, P., Eggli, Y., Pretre-Rohrbach, I., Meylan, D., Marzzi, A., & Burnand, B. (2006). Validation of the Potentially Avoidable Hospital Readmission Rate as a Routine Indicator of Quality of Care. *Medical Care*, 44(11), 972-981.

Helling, D.K., Nelson, K.M., Ramirez, J.E., & Humphries, T.L. (2006). Kaiser Permanente Colorado Region Pharmacy Department: innovative leader in pharmacy practice. *Journal of the American Pharmacists Association*, 46, 67–76.

Hofmarcher, M.M., Paterson, I., & Riedel, M. (2002). Measuring hospital efficiency in Austria – a DEA approach. *Health Care Management Science*, 5, 7–14.

Hollingsworth, B. (2003). Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science,* 6(4), 203–218.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, 17(10), 1107–1128.

Institute of Medicine. (2000). *To Err Is Human: Building a Safer Health System*. PubMed. https://pubmed.ncbi.nlm.nih.gov/25077248/

Jacobs, R. (2001). Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. *Health Care Management Science*, 4(2), 103–115.

Jacobs, R., Goddard, M., & Smith, P. C. (2005). How Robust Are Hospital Ranks Based on Composite Performance Measures? *Medical Care*, 43(12), 1177–1184.

Jacobs, R., Smith, P. C., & Street, A. (2006). *Measuring Efficiency in Health Care: Analytic Techniques and Health Policy*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511617492.

Jia, T., & Yuan, H. (2017). The application of DEA (Data Envelopment Analysis) window analysis in the assessment of influence on operational efficiencies after the establishment of branched hospitals. *BMC Health Services Research*, 17(1), 265-273.

Karagiannis, R., & Karagiannis, G. (2018). Intra and inter-group composite indicators using the BoD model. *Socio-Economic Planning Sciences*, 61, 44–51.

Khezrimotlagh, D., Cook, W. D., & Zhu, J. (2021). Number of performance measures versus number of decision making units in DEA. *Annals of Operations Research*, 303(1), 529-562.

Khushalani, J., & Ozcan, Y. (2017). Are hospitals producing quality care efficiently? An analysis using Dynamic Network Data Envelopment Analysis (DEA). *Socio-Economic Planning Sciences*, 60, 15-23.

Kleinsorge, I. K., & Karney, D. F. (1992). Management of nursing homes using data envelopment analysis. *Socio-economic Planning Sciences*, 26(1), 57–71.

Kohl, S., Schoenfelder, J., Fügener, A., & Brunner, J. O. (2019). The use of data envelopment analysis (DEA) in healthcare with a focus on hospitals. *Health Care Management Science*, 22(2), 245–286.

Kontodimopoulos, N., Nanos, P., & Niakas, D. (2006). Balancing efficiency of health services and equity of access in remote areas in Greece. *Health Policy*, 76(1), 49–57.

Kooreman, P. (1994). Nursing home care in the Netherlands: A nonparametric efficiency analysis. *Journal of Health Economics*, 13(3), 301-316.

Laine, J., Finne-Soveri, U. H., Björkgren, M., Linna, M., Noro, A., & Häkkinen, U. (2005a). The association between quality of care and technical efficiency in long-term care. *International Journal of Quality in Health Care,* 17(3), 259-267.

Laine, J., Linna, M., Häkkinen, U., & Noro, A. (2005b). Measuring the productive efficiency and clinical quality of institutional long-term care for the elderly. *Health Economics*, 14, 245–256.

Lee, K., Lee, J., & Chun, K. (2008). Reforming the hospital service structure to improve efficiency: Urban hospital specialization. *Health Policy*, 87(1), 41–49.

Lin, C.S., Chiu, C.M., Huang, Y.C., Lang, H.C., & Chen, M.S. (2022). Evaluating the Operational Efficiency and Quality of Tertiary Hospitals in Taiwan: The Application of the EBITDA Indicator to the DEA Method and TOBIT Regression. *Healthcare*, 10(1), 58.

Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., & Bratzler, D. W. (2007). Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine*, 356(5), 486-496.

Linna, M. (1997). Measuring hospital cost efficiency with panel data models. *Health Economics,* 7(5), 415–427.

Linna, M., Häkkinen, U., & Magnussen, J. (2006). Comparing hospital cost efficiency between Norway and Finland. *Health Policy,* 77(3), 268–278.

Madureira, C. (2020). A reforma da administração pública e a evolução do estado-providencia em Portugal: história recente. *Ler História*, 76(1), 179–202.

Magnussen, J. (1996). Efficiency measurement and the operationalization of hospital production*. Health Services Research,* 31(1), 21-37.

Marahem, M., Ghodraty, M. S., Farzin, H., Afamohammadi, D., & Erçelen, O. (2018). The importance of bed occupancy rate in evaluating the utilization of hospitals. *Crescent Journal of Medical and Biological Sciences,* 4(5), 270-271.

Mark, B. A., Jones, C. B., Lindley, L., & Ozcan, Y. A. (2009). An examination of technical efficiency, quality, and patient safety in acute care nursing units. *Policy Politics Nursing Practice*, 10(3), 180-186.

Matos, R. (2020). *Análise económico-financeira dos hospitais EPE portugueses*. [Unpublished master's dissertation]. Department of Engineering and Management, Instituto Superior Técnico.

Matos, R., Ferreira, D. C., & Pedro, I. (2021). Economic analysis of Portuguese public hospitals through the construction of quality, efficiency, access, and financial-related composite indicators. *Social Indicators Research*, 157(1), 361–392.

Mattke, S., Kelley, E., Scherer, P., Hurst, J., & Lapetra, M. L. G. (2006). *Health care quality indicators project initial indicators report.* OECD Health Working Papers. https://doi.org/10.1787/481685177056

Maxwell, R. J. (1992). Dimensions of quality revisited: From thought to action. *Quality in Health Care*, 1(1), 171–177.

McNair, C. J., & Leibfried, K. H. J. (1992*). Benchmarking: A Tool for Continuous Improvement*. Omneo. https://books.google.pt/books/about/Benchmarking.html?id=ZJ8oAQAAMAAJ&redir_esc=y.

Miot, H. A. (2018). Análise de correlação em estudos clínicos e experimentais. *Jornal Vascular Brasileiro,* 17(4), 275-279.

Miranda, R. O. D. (2009*). Eficiência técnica de cursos de administração: uma análise conjunta de DEA e SFA*. [Unpublished master's dissertation]. Instituto de Ensino e Pesquisa.

Mitropoulos, P., Talias, M. A., & Mitropoulos, I. (2015). Combining stochastic DEA with Bayesian analysis to obtain statistical properties of the efficiency scores: An application to Greek public hospitals. *European Journal of Operational Research*, 243(1), 302-311.

Morais, P., & Camanho, A. S. (2011). Evaluation of performance of European cities with the aim to promote quality of life improvements. *Omega*, 39(4), 398-409.

Moreira, S. (2008). *Análise da Eficiência dos Hospitais-Empresa: Uma Aplicação da Data Envelopment Analysis* | Banco de Portugal. https://www.bportugal.pt/paper/analise-da-eficiencia-dos-hospitais-empresa-uma-aplicacao-da-data-envelopment-analysis.

Mosel, D., & Gift, B. (1994). Collaborative benchmarking in health care. The Joint Commission Journal on Quality Improvement, 20(5), 239-249.

Mota, J.C. (2016). *Cirurgia Ambulatória: Readmissões pós-cirúrgicas* [Unpublished master's dissertation]. Escola Nacional de Saúde Pública, ENSP.

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). *Tools for Composite Indicators Building.* JRC Publications Repository. https://publications.jrc.ec.europa.eu/repository/handle/JRC31473.

Nardo, M., Saisana, M., Saltelli, A., Giovannini, E., Tarantola, S., & Hoffman, A. (2008). *Handbook on Constructing Composite Indicators and User Guide.* OECD; European Commission, Joint Research Centre. https://doi.org/10.1787/533411815016.

Navarro-Espigares, J.L. and Torres, E.H. (2011). Efficiency and quality in health services: a crucial link. *The Service Industries Journal,* 31(3), 385-403.

Nayar, P., & Ozcan, Y. (2008). Data envelopment analysis comparison of hospital efficiency and quality. *Journal of Medical Systems*, 32(3), 193-199.

Nayar, P., Ozcan, Y. A., Yu, F., & Nguyen, A. T. (2013). Benchmarking urban acute care hospitals: Efficiency and quality perspectives. *Health Care Management Review*, 38(2), 137–145.

Nunamaker, T.R. (1983). Measuring Routine Nursing Service Efficiency: A Comparison of Cost per Patient Day and Data Envelopment Analysis Models. *Health Services Research*, 18(2), 183-205.

Nunes, A. S. R. A. de M. M. (2019). *Eficiência e produtividade nas unidades hospitalares em Portugal: análise comparada dos regimes PPP e E.P.E.* [Unpublished masther's dissertation]. Universidade de Lisboa.

Nunes, A. & Matos, A. (2017). Experiências de Gestão Hospitalar Realizadas em Portugal - Uma Perspectiva Comparada. *Revista Eletrônica Gestão e Saúde*, 8(2), 284-301.

Nunes, A. M., & Ferreira, D. C. (2022a). A Critical Analysis of Decentralizing the Portuguese Public Healthcare Provision Services. *International Journal of Environmental Research and Public Health*, 19(20), 13390.

Nunes, A. M., & Ferreira, D.C. (2022b). Evaluating Portuguese Public Hospitals Performance: Any Difference before and during COVID-19? *Sustainability,* 15(1), 294.

Nunes, A. M., & Ferreira, D. C. (2022c). Social Inequity and Health: From the Environment to the Access to Healthcare in Composite Indicators, the Portuguese Case. In *Sustainable Policies and Practices in Energy, Environment and Health Research* (pp. 371-389). Springer. https://doi.org/10.1007/978-3-030-86304-3_21

Nunes, A.M., & Ferreira, D. (2018). The health care reform in Portugal: outcomes from both the New Public Management and the economic crisis. *The International Journal of Health Planning and Management*, 33(3), 1-20.

Nunes, R. (2014). *Regulação da saúde* (3ª ed). [Health Regulation]. Porto: Vida Económica.

O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences*, 42(3), 158–189.

OECD (2008). *Handbook on constructing composite indicators: methodology and user guide*. European Commission, Joint Research Centre and OECD. https://doi.org/10.1787/9789264043466-en.

Oliveira, R. C., & Ferreira, J. S. (2014). *Investigação operacional em ação: casos de aplicação* (1ª ed.) [Operational investigation in action: application cases]. Ensino. https://doi.org/10.14195/978-989-26-0738-2.

Ozcan, Y. A. (2008). *Health Care Benchmarking and Performance Evaluation: An Assessment using Data Envelopment Analysis (DEA)*. Springer. https://doi.org/10.1007/978-0-387-75448-2.

Ozcan, Y. A. (2009). Quantitative Methods in Health Care Management: Techniques and Applications (2ª ed.). Jossey-Bass https://download.e-bookshelf.de/download/0000/5729/42/L-G-0000572942-0002358371.pdf

Pereira, M. A., Camanho, A. S., Figueira, J. R., & Marques, R. C. (2021a). Incorporating preference information in a range directional composite indicator: The case of Portuguese public hospitals. *European Journal of Operational Research*, 294(2), 633–650.

Pereira, M. A., Ferreira, D. C., Figueira, J. R., & Marques, R. C. (2021b). Measuring the efficiency of the Portuguese public hospitals: A value modeled network data envelopment analysis with simulation. *Expert Systems with Applications*, 181(1), 115169.

Portela, M. C. A. S., Camanho, A. S., Almeida, D. Q., Lopes, L., Silva, S. N., & Castro, R. (2016). Benchmarking hospitals through a web-based platform. *Benchmarking: An International Journal*, 23(3), 722-739.

Puyenbroeck, T. V. (2017). On the output orientation of benefit-of-doubt model. *Social Indicators Research*, 139, 415–431.

Rocha, A., Costa, A. S., Figueira, J. R., Ferreira, D. C., & Marques, R. C. (2021). Quality assessment of the Portuguese public hospitals: A multiple criteria approach. *Omega*, 105, 102505.

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A*, 168(2), 307–323.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer.* John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470725184.

Sarkis, J. (2007*). Preparing your data for DEA. In Modeling data irregularities and structural complexities in data envelopment analysis*. Springer. https://doi.org/10.1007/978-0-387-71607-7_17.

Sherman, H. D. (1984). Hospital efficiency measurement and evaluation: Empirical test of a new technique. *Medical Care,* 9, 922–938.

Shwartz, M., Burgess, J. F., & Berlowitz, D. (2009). Benefit-of-the-doubt approaches for calculating a composite measure of quality. *Health Services and Outcomes Research Methodology*, 9(4), 234–251.

Silva, C., Silva, S., & Gonçalves, C. (2020). Data Envelopment Analysis (DEA) em Estudos sobre Saúde e Educação: uma Análise Bibliométrica nos últimos 30 anos. *Revista FSA*, 18(1), 214-239.

Simões, P., & Marques, R. (2011). Performance and congestion analysis of the Portuguese hospital services. *Central European Journal of Operations Research*, 19(1), 39-63.

Singaroyan, R., Seed, C. A., & Egdell, R. M. (2006). Is a target culture in health care always compatible with efficient use of resources? A cost-effectiveness analysis of an intervention to achieve thrombolysis targets. *Journal of Public Health*, 28(1), 31–34.

Soares de Mello, J. C. C. B., Meza, L. A., Gomes, E. G., Serapião, B. P., & Lins, M. P. E. (2003). Análise de envoltória de dados no estudo da eficiência e dos benchmarks para companhias aéreas brasileiras. *Pesquisa Operacional*, 23(2), 325–345.

Stevanovic, V., Feek, C., & Kay, R. (2005). Using routine data for benchmarking and performance measurement of public hospitals in New Zealand. *Benchmarking: An International Journal*, 12(6), 498-507.

Thomas, J.W. (1996). Does risk-adjusted readmission rate provide valid information on hospital quality? *Inquiry*, 33(3), 258-270.

Thomas, J.W., & Holloway, J.J. (1991). Investigating early readmission as an indicator for quality of care studies. *Medical Care*, 29(4), 377-394.

Thompson, R.G., Singleton, F.D., Thrall, R.M., & Smith, B. A. (1986). Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas. *Interfaces*, 16(6), 35–49.

Vara, G., Gomes, M. C., & Ferreira, D. C. (2023). *Optimistic and Pessimistic Benchmarking Approaches for Assessing the Performance of Healthcare Entities in Pandemic Times: The Case of Portuguese Public Hospitals Before, During, and After COVID-19 Outbreak*. Social Science Research Network. https://doi.org/10.2139/ssrn.4401834

Varanda, J., Gonçalves, L., & Craveiro, I. (2020). The Unlikely Saviour: Portugal's National Health System and the Initial Impact of the COVID-19 Pandemic. *Development*, 63(4), 291–297.

Vasquez, M. C. (2012). *Eficiência e produtividade no ensino superior público*. [Unpublished doctoral' dissertation]. Instituto de Ciências Sociais e Políticas, Universidade de Lisboa.

Vidoli, F., Fusco, E., Pignataro, G., & Guccio, C. (2023). *Multi-directional Robust Benefit of the Doubt model: A comprehensive measure for the quality of health care in OECD countries.* Health, Econometrics and Data Group (HEDG) Working Papers (23/14). https://ideas.repec.org/p/yor/hectdg/23-14.html.

Werner, R. M., & Bradlow, E. T. (2006). Relationship Between Medicare's Hospital Compare Performance Measures and Mortality Rates. *JAMA*, 296(22), 2694.

WHO. (2006). *Quality of care: a process for making strategic choices in health systems.* Pesquisa.bvsalud.org. https://pesquisa.bvsalud.org/portal/resource/pt/mis-21350

Worthington, A. C. (2004). Frontier efficiency measurement in health care: A review of empirical studies. *Medical Care Research and Review*, 61(2), 135–170.

Xiaoya, L., & Jinchuan, C. (2008). A comprehensive DEA approach for the resource allocation problem based on scale economies classification. *Journal of Systems Science and Complexity*, 21(4), 540-557.

Zhan, C., & Miller, M. R. (2003). Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. *JAMA*, 290(14), 1868–1874.

Zhou, P., Ang, B., & Poh, K. (2007). A mathematical programming approach to constructing composite indicators. *Ecological Economics*, 62(2), 291–297.

Appendix

Table 8- Literature review (Source: Author).

Authors	Case Study	Sample	Model(s)	Variables	Conclusions
Grosskopf &Valdmanis (1987)	This study assesses the relative performance of a sample of hospitals in California. Hospitals are compared on the basis of their relative technical efficiency.	22 public and 60 private, not-for-profit hospitals.	DEA; SFA; COLS;	Inputs: FTE physicians, FTE non-physicians, Net Plant Assets (NPA), Doctors per 1000 inhabitants, Beds, Average length of hospital stay in ICU (intensive care units), Average length of hospital stay in acute care units. Outputs: Number of surgeries OP, Number of surgeries IP, Number of outpatient visits + emergency room visits, Number of admissions.	Public hospitals in are less resource using than NFP hospitals, i.e., for a given output mix and level they would have lower costs. The results also demonstrate that NFP hospitals provide higher quality care if relatively intense use of inputs in general, and non-physician labor in particular, is considered to be correlated with the indicators of quality.
Burgess & Wilson (1995)	Measure technical efficiency among U.S. hospitals from 1985-1988, which are evaluated by ownership type--profit, nonprofit, state and local government, and U.S. Department of Veterans Affairs (VA).	137 nonpsychiatric hospitals in the VA system.	Malmquist index; DEA;	Inputs: Beds, Beds long-term, FTE Registered licensed practical nurses, FTE Other staff, FTE non-physician, FTE physicians (long-term care), Average length of hospital stay. Outputs: Average length of hospital stay in ICU (intensive care units), Case-mix weighted inpatient discharges, Number of outpatients, Number of surgeries OP, Number of surgeries IP.	Between 1985 and 1988, hospitals experienced a one-time effect on efficiency due to industrial developments in the healthcare sector. Technology changes led to increased healthcare expenditure, not inefficiency. However, a net technical regress suggests a decline in productivity, resulting in higher costs for patient care due to increased labor and bed-days.
Magnussen (1996)	To discuss the usefulness of efficiency measures as instruments of monitoring and resource allocation by analyzing their invariance to changes in the operationalization of hospital production.	46 Norwegian acute care, non-teaching hospitals.	DEA;	Inputs: Beds, FTE physicians, FTE Other staff, FTE Registered nurses, Average length of hospital stay in long-term care. Outputs: Number of outpatients, Number of surgery patients, Number of medical patients.	The distribution of efficiency is found to be unaffected by changes in the specification of hospital output. Both the ranking of hospitals and the scale properties of the technology, however, are found to depend on the choice of output specification.
Dismuke & Sena (1999)	Assess the impact of actual DRG payment on the productivity (through its components, i.e., technological change and technical efficiency change) of diagnostic technology in Portuguese hospitals.	48122 discharges for DRG 14 and 29021 for DRG 127.	SFA; DEA; Malmquist index;	Outputs: Number of patients discharged, Number of CAT scans for DRG 14, Number of EEG for DRG 127 used in producing discharges, Number of ECO for DRG 127 used in producing discharges.	The DRG payment system positively impacts productivity and technical efficiency of diagnostic technologies in Portugal, according to a study. The authors recommend the Health Ministry continue implementing the DRG-based payment system to enhance healthcare system effectiveness and address shifts in productivity among the three diagnostic technologies studied.

Hofmarcher et al. (2002)	Investigate the evolution of efficiency and productivity in the hospital sector of an Austrian province for the years 1994-1996.	31 wards in Austria.	DEA;	Inputs: Beds, FTE physicians, Paramedical personnel, Administrative staff, LDF-points, Average length of hospital stay. Outputs: Number of patients discharged.	The Austrian hospital industry's efficiency levels were analyzed using two models. The first model showed efficiency levels exceeding 95%, indicating that less than 5% of inputs could be conserved while maintaining output. The second model showed a noticeable decline in efficiency when using LDF-points. Overall, efficiency changes between 1994 and 1996 were minimal.
Carey (2003)	Estimate a stochastic frontier cost function to test for inefficiency differences among system hospitals having common strategic and/or structural characteristics.	1209 US hospitals.	SFA;	Inputs: Average annual salary, Beds, Average length of hospital stay. Outputs: Number of inpatient admissions, Case-mix index, Mortality rate.	The study reveals that hospital affiliation with multi-hospital systems is controversial and has conflicting evidence about the cost of system hospitals. It suggests that system hospitals with coordinated physician and insurance networks and decentralized service delivery are less inefficient. Centralized physician integration and insurance product development may also be beneficial. The study suggests that hospital systems should be defined based on strategic and structural attributes.
Grosskopf et al. (2004)	Data envelopment analysis (DEA) approach to measure the relative technical and scale efficiencies on a sample of 254 US teaching hospitals.	236 teaching hospitals and 556 non-teaching hospitals in the US.	DEA;	Inputs: Beds, FTE physicians with staffing privileges, FTE Registered nurses, FTE Registered licensed practical nurses, FTE medical residents, Other staff. Outputs: Number of inpatients, Number of inpatients + outpatient surgeries, Number of outpatient visits + emergency room visits.	The study reveals that competition positively impacts teaching hospitals, enhancing efficiency, instructional commitment, and intensity. Additionally, commitment to teaching, non-public ownership, and beneficial affiliation effects improve productivity. However, association with a medical school has minimal impact on output.
Jacobs et al. (2005)	The study examines the use of composite indices in healthcare performance, highlighting how data variability can undermine their robustness and how methodological rules can significantly impact composite scores.	117 English acute hospitals.	Stochastic modeling;	Inputs: Junior doctors' working hours, Data quality for National Health, Service trusts with inpatient activity. Outputs: Mortality within 30 days of surgery, Readmission within 28d of discharge after treatment for fractured hip, Number of patients discharged after fractured hip, Percentage of patients waiting less than 6 months for admission, Number of outpatient visits with waiting time of 13 weeks, Sickness absence rate, Satisfaction with healthcare, Staff opinion survey.	Random variation, which is beyond the control of organizations, introduces considerable uncertainty into hospital scores. Furthermore, the results of composite indicators are sensitive to changes in the weighting system and aggregation rules. As a result of subtle changes, some hospitals can move up or back by almost half of the league table.

Laine et al. (2005a)	To analyse the association between quality of care and technical (productive) efficiency in institutional long-term care wards for the elderly.	114 wards in 41 facilities in Finland.	DEA;	Inputs: FTE Registered nurses, FTE Registered licensed practical nurses, Aides average number, Beds. Outputs: Case-mix weighted inpatient days.	The study found that wards with high physical restrictions or bedfast residents outperformed the comparator group. The findings suggest a correlation between technical effectiveness and undesirable quality aspects. Therefore, elder care management and performance evaluation should consider efficiency and care quality.
Laine et al. (2005b)	Applied the technical inefficiency effect estimation method proposed by Battese and Coelli to explore the effect of selected quality measures on efficiency.	122 wards in health-centre hospitals and residential homes in Finland.	SFA; DEA;	Inputs: FTE Registered licensed practical nurses, FTE Registered nurses, FTE person years for foi aides, Beds. Outputs: Prevalence of pressure ulcers, Prevalence of weekly use of depressants and hypnotics, Prevalence of depression with no treatment, Case-mix weighted inpatient days	Three quality variables did not show a correlation with inefficiency, but low-pressure ulcer prevalence was linked to efficiency. Health center-hospitals were more efficient, and pressure ulcers were key performance indicators for management. Quality measures based on depression medication and treatment positively correlated with efficiency.
Stevanovic et al. (2005)	This paper aims to outline benchmarking and performance monitoring practice developed by the New Zealand Health Information Service at a diagnosis-related group (DRG) level.	40 major hospitals in New Zealand.	CBD; Benchmarkin g;	Outputs: Average length of hospital stay, Number of day hospital sessions, Number of high length-of-stay outliers, Number of discharge transfers to another facility, Readmission rate of major diagnostic category (MDC) related same-hospital, Mortality rate, Average patient clinical complexity level (PCCL) per record, Average number of MDCs per record.	The study discusses the use of benchmarking and performance assessment in healthcare to improve service standards, reduce costs, and enhance patient outcomes. It emphasizes the benefits of routine data, including standardization, timeliness, completeness, and sample size, and the potential for in-depth research.
Kontodimopo ulos et al. (2006)	Investigate the efficiency of a set of small-scaled Greek hospitals known as hospital health centers, obtaining insight on their productive efficiency in light of their particular role.	17 small-scaled Greek hospitals.	DEA;	Inputs: Doctors per 1000 inhabitants, Nurses per 1000 inhabitants, Beds occupancy rate, Beds, Average length of hospital stay. Outputs: Number of inpatient admissions, Number of outpatients, Preventive medicine services, Number of patient days.	The study reveals significant operational inefficiencies in Greek hospitals and rural healthcare facilities, requiring reduction while maintaining equal healthcare access. Two potential solutions are telemedicine and enhancing preventative medical services. The results show technical inefficiencies of 26.77% and 25.13%, respectively. Remote units, particularly on small islands, are more inefficient. However, units offering preventive services perform better.
Linna et al. (2006)	Comparison of cost efficiency in Finnish and Norwegian hospitals using national discharge data and identical definitions for cost and output measures.	47 Finnish and 51 Norwegian public hospitals in 1999.	DEA;	Inputs: Operational expenses, Average length of hospital stay. Outputs: DRG weighted admissions, Number of outpatients, Number of inpatients (short-stay).	The study reveals significant cost efficiency discrepancies between Norwegian and Finnish hospitals, partly due to variations in input costs and lengths of stay. It emphasizes the importance of accurate input pricing indices in cross-country comparisons. Future studies should include financial factors and a wider range of observable features. The study suggests that global hospital benchmarking can identify effective facilities and accelerate technology advances, but performance indicators must be comparable.

Werner & Bradlow (2006)	Determine whether quality measured with the process measures used in the CMS's Hospital Compare are correlated with and predictive of hospitals' risk-adjusted mortality rates.	3657 US hospitals.	Bayesian Approach;	Inputs: Percentage of AMI patients who received aspirin within 24 h of arrival, Percentage of AMI patients who received beta blocker within 24 h of arrival, Percentage of CHF patients who received an ACE inhibitor for left ventricular dysfunction., Percentage of CHF patients who received an assessment of left ventricular function, Percentage of pneumonia patients' timing of initial antibiotic therapy, Percentage of pneumonia patients who receive pneumococcal vaccination, Percentage of pneumonia patients who receive oxygenation assessment within 24h of admission. Outputs: Percentage of AMI patients prescribed beta blocker at discharge, Percentage of AMI patients with aspirin therapy prescribed at discharge.	The study suggests that performance measurements can be used to link performance to rewards, with incentive programs more successful if they focus on the lowest performing hospitals. However, the study may have flaws, such as potential confounding variables and unreported patient and hospital characteristics. The risk-adjustment models, based on administrative data, may have been inaccurate due to underreported comorbidities. The study suggests that establishing health care quality metrics more closely connected to patient outcomes is crucial for improving healthcare quality.
Arocena & Garcia-Prado (2007)	The paper explores how Costa Rican public hospitals responded to 1997-2001 reforms, after the introduction of management contracts, by constructing a productivity index that accounts for productivity changes while maintaining quality of care.	20 public hospitals in Costa Rica.	Output distance function; Malmquist index;	Inputs: FTE physicians, FTE Registered nurses, Expenses with supplies and external services per severity-adjusted patient, Beds, Average length of hospital stay. Outputs: Readmissions within 30 days after discharge per 100 inpatients, Absence rate among health care personnel, Intra-hospitals infection rate, Mortality rate, Number of patients discharged, Number of outpatients.	According to the authors' findings, quality improvements are the key factor improving hospital performance. This improvement appears to have been facilitated by the use of management contracts, particularly for small hospitals. Additionally, technical and scale efficiency changes rather than technology changes are the main causes of productivity development.
Barros et al. (2007)	Analyse of the efficiency and productivity growth of a representative sample of Portuguese hospitals from 1997 to 2004, using an innovative approach by employing the directional distance function and the Luenberger productivity indicator.	51 hospitals in Portugal.	Luenberger productivity indicator; directional distance function;	Inputs: Beds, Staff and administrative personnel, Total cost, Average length of hospital stay. Outputs: Number of patients discharged, Number of emergency episodes, Number of medical appointments.	The study found that productivity increase in Portuguese hospitals between 1997 and 2004 was not observed in facilities with no technical efficiency or technological advancements. The study suggests that capital accumulation can boost both efficiency and technology. Future research should consider different time periods and variables to compare methodologies and identify investment episodes and policy reforms related to productivity.
Giraldes (2007)	Evaluate, from an efficiency and quality perspective, the Hospitals and compare them with the results of the evaluation previously carried out in September 2005.	6 groups of EPE and SPA Hospitals.	Quality Assessment Aggregate Indicator;	Inputs: Total operating expenses per bed, Expenses with external consultations, Expenses with urgencies per patient, Expenses with medicine per standard patient, Expenses with medicine in day hospital per standard patient, Expenses with medicine per external consultation, Expenses with medicine in urgency per standard patient, Expenses with auxiliary clinical support sections, Expenses with medical-surgical supplies. Outputs: Number of surgeries OP, Number of caesarean sections per 100 deliveries, Autopsy rate, Percentage of inpatient admissions for surgical infections in the total number of surgical procedures.	The study reveals significant variations in hospital efficiency and quality in Portugal, with hospitals like Hospital de São Sebastião and Hospital de Aveiro being the least efficient. Central hospitals like CH Vila Real/Régua show the best efficiency. EPE hospitals generally outperform SPA hospitals in terms of efficiency. Inequalities include expenses, medicines, outpatient consultations, and physical medicine and rehabilitation per patient treated.

Lindenauer et al. (2007)	Measure changes in adherence to 10 individual and 4 composite measures of quality over a period of 2 years at 613 hospitals that voluntarily reported information about the quality of care through a national public-reporting initiative. Comparison of the pay-for-performance hospitals with the 406 hospitals with public reporting only (control hospitals).	4691 hospitals that submitted data for the HQA between the fourth quarter of 2003 and the third quarter of 2005.	Multivariable modeling;	Inputs: Percentage of AMI patients who received aspirin within 24 h of arrival, Percentage of AMI patients with ACE inhibitor for left ventricular dysfunction, Percentage of AMI patients who received beta blocker within 24 h of arrival, Percentage of CHF patients who received an assessment of left ventricular function, Percentage of CHF patients who received an ACE inhibitor for left ventricular dysfunction, Percentage of pneumonia patients who receive oxygenation assessment, Percentage of pneumonia patients who receive first antibiotics within 4 h of arrival, Percentage of pneumonia patients who receive pneumococcal vaccination. Outputs: Percentage of AMI patients with aspirin therapy prescribed at discharge, Percentage of AMI patients prescribed beta blocker at discharge.	Hospitals participating in pay-for-performance programs showed significant improvements in all quality composite measures compared to the control group, with greater improvements observed in hospitals with lower baseline performance. After adjusting for baseline performance and other hospital characteristics, pay for performance led to improvements ranging from 2.6% to 4.1% over two years. Hospitals engaged in both public reporting and pay for performance achieved slightly better quality improvements compared to those involved only in public reporting. Further research is needed to explore alternative incentives and assess the overall cost-effectiveness of these programs.
Afonso & Fernandes (2008)	Estimate in a non-parametrical framework measures of hospital technical efficiency during the period 2000-2005. Additionally, examine heterogeneity and efficiency resulting from the recent "privatisation" of some Portuguese hospitals.	68 Portuguese public hospitals.	DEA; Malmquist indexes;	Inputs: Doctors per 1000 inhabitants, Nurses per 1000 inhabitants, Other staff, Beds, Average length of hospital stay. Outputs: Weighted number of patients who leave hospital after inpatient admission, Number of outpatients, Number of emergency episodes, Number of surgeries produced, Possession of CAT equipment.	The NHS hospital sector experienced small productivity growth between 2000 and 2004, with a slight decrease in 2005. Efficiency scores fluctuated significantly, with only 7 hospitals remaining on the production possibility frontier. 15% of hospitals' efficiency scores fell between 2000 and 2005, indicating potential difficulties in productivity over time and room for improvement.
Lee et al. (2008)	This study examines the relationship between the case-mix specialization index and inpatient hospital care service efficiency in Seoul, Korea. It uses information theory index and data envelopment analysis to identify internal and external factors affecting specialization status.	106 acute care hospitals in Seoul.	DEA; Multiple regression analysis;	Inputs: Beds, Nurses per 1000 inhabitants, Doctors per 1000 inhabitants. Outputs: Number of inpatients, Number of outpatients.	The study reveals that only 11.3% of Seoul's 106 hospitals are efficient, with teaching hospitals having higher efficiency scores. Organizational attributes explain provider specialization status more than market conditions. Specialization can improve hospital operations, and converting into specialized hospitals can increase competitiveness and financial stability.
Moreira (2008)	Evaluation of the impacts of the 2022 reform of the public hospital sector on technical efficiency, through the analysis of the relative performance of company-hospitals, comparatively to hospitals that remained within the public administration sector, before and after the entrepreneurialization process.	64 public hospitals.	DEA;	Inputs: Beds, Doctors per 1000 inhabitants, Nurses per 1000 inhabitants, Paramedical personnel, Other staff, Total staff, Expenses with materials consumed, Expenses with external supplies and services, Expenses with staff compensation, Total cost. Outputs: Number of patients discharged, Number of external consultations, Number of day hospital sessions, Number of emergency appointments, Number of surgeries OP.	The study used global border analysis and group border analysis to compare efficiency gains in corporate hospitals compared to public administrations. Results showed efficiency gains, but not significant. The worst start-up situation may limit improvement opportunities. The study suggests increasing diversity and quality of information, especially in a corporatized context, to improve efficiency.

Nayar & Ozcan (2008)	Use DEA models to compare hospital performance in terms of technical efficiency and quality, and to determine whether there is an efficiency–quality trade-off.	117 non-federal acute care hospitals in Virginia with non-missing data in the AHA survey database.	DEA;	Inputs: Beds, Expenses with supplies and external services per severity-adjusted patient, Total staff, Total assets. Outputs: Number of patients discharged, Number of outpatients, Training FTE.	The study reveals that improving technical efficiency doesn't compromise quality, a crucial factor for policymakers and administrators. Hospitals can maximize both quality and quantity, and quality measures in DEA can serve as a benchmarking tool. The study found a high correlation between efficiency scores obtained using only technical outputs and efficiency scores obtained using technical combined with quality outputs.
Mark et al. (2009)	Assess the technical efficiency of general medical, general surgical, and combined medical–surgical nursing units, incorporating into the DEA model several relevant indicators of quality and patient safety.	226 units in 118 hospitals.	DEA;	Inputs: Nurses working time, Licensed practical nurses working time, Unlicensed assistive personnel hours of care, Operational expenses, Beds. Outputs: Number of unit discharges, case mix adjusted, Satisfaction with healthcare, Number of medication errors per 1000 patient days, Number of patient falls per 1000 patient days.	The study revealed that 60% of nursing units were inefficient, with traditional efficiency targets not yielding significant gains. Key improvements include reducing labor hours and medication errors. The study emphasizes the importance of improving patient safety for improving nursing unit efficiency, providing insights into structural and payment alternatives.
Shwartz et al. (2009)	The study explores two BoD approaches, LP models and DEA, for assessing facility rankings in VA nursing homes. Composite scores from 32 VA nursing homes are calculated using standard and BoD approaches and the tradeoff between weight adjustment and rankings is also examined.	32 Department of Veterans Affairs (VA) nursing homes.	DEA; BoD;	Outputs: Prevalence of pressure ulcers, Functional decline, Behavioral decline, Mortality rate within 6 months of discharge, Readmission rate within 6 months after discharge.	BoD methods like equal weighting and expert opinion acknowledge local performance variations, allowing facilities to modify weights and facilitating dialogue. The DEA approach offers transparency, combining quality indicators without normalizing individual scores, and identifying benchmark facilities.
Blank & Valdmanis (2010)	The paper discusses the efficiency of Dutch hospitals using Data Envelopment Analysis (DEA), focusing on cost inefficiency measures due to their operating environment. The authors use Simar and Wilson bootstrapping techniques to obtain more efficient estimates of environmental effects.	69 Dutch hospitals.	DEA; bootstrapping;	Inputs: Staff and administrative personnel, Nurses per 1000 inhabitants, Paramedical personnel, Other staff, Material supplies, Expenses with staff and administrative personnel, Expenses with nursing personnel, Expenses with paramedical personnel, Expenses with staff compensation, Expenses with external supplies and services, Variable cost. Outputs: Number of first-time visits, Number of unit discharges, case mix adjusted.	The study concludes that physician intensity greatly increases cost inefficiency, presumably as a result of doctors substituting hospital inputs for their own time. Neglecting how doctors' performance affects hospitals may result in reduced payments, which might force hospitals to cut back on services because of financial restrictions. Management inefficiency differs amongst hospitals even after accounting for environmental conditions.
Garcia-Lacalle & Martin (2010)	The study examines the impact of market-driven reforms on efficiency and quality in urban and rural hospitals in the European health service, comparing their performance in the Andalusian Health Service (SAS), in Spain.	27 Andalusian Health Service (SAS) hospitals in Spain.	DEA;	Inputs: Beds, FTE Registered nurses, FTE physicians. Outputs: Number of inpatients, Number of surgeries produced number of medical diagnoses, Satisfaction with healthcare, Number of outpatients, Number of emergency episodes.	Results show that rural hospitals perform similarly in efficiency, but significantly better in patient satisfaction. Some rural hospitals are the best performers, suggesting that market-driven reforms do not necessarily affect the performance of both urban and rural hospitals. The findings indicate that quality and efficiency may coexist as goals.

Navarro-Espigares & Torres (2011)	This paper examines the evolution of efficiency and quality in Andalusian hospitals from 1997-2004, using data envelopment analysis and the Malmquist productivity index, and also studies the association between efficiency and quality indicators.	27 Andalusian Health Service hospitals.	DEA;	Inputs: FTE physicians, FTE Registered nurses, FTE Other staff, Beds, Average length of hospital stay. Outputs: Ambulatory weighted health care units, Diagnostic related group, Preoperative average stay, Case-mix index, Stays utilization index, Percentage of scheduled surgeries over the total carried out, Surgical turnover, Number of caesarean sections per 100 deliveries, Average number of diagnoses and procedures at discharge, Number of procedures at discharge, Satisfaction with healthcare.	The study on the technical efficiency of Andalusian Health Service hospitals from 1997 to 2004 showed an increasing trend, with almost all variables exhibiting a positive evolution. The study demonstrated that there was no trade-off between efficiency and quality goals, and that output variables that reflect both the activity undertaken and the level of services must be included in the overall assessment of health system performance.
Barros et al. (2013)	This article aims to investigate the statistical factors that characterize the hospitals, clustering the sample in segments that enable to define specific health policies for identified segments and to monitor each group with adequate targets.	51 hospitals in Portugal.	Stochastic cost frontier;	Inputs: Mix of specialist services, Beds, Expenses with staff compensation, Expenses with of other factors proxied by the regional price index, Average length of hospital stay. Outputs: Number of patients discharged, Number of external consultations, Number of emergency appointments.	The study reveals that Portuguese hospitals exhibit distinct levels of productivity and sustainability, suggesting that a uniform government strategy may not be suitable for all. It identifies three distinct hospital segments with varying features, performance levels, and debts, with private hospitals being more effective than governmental hospitals. Good hospitals operate due to a desire for quality, while those lacking quality systems use them to cover up poor managerial practices.
Ferrier & Trivitt (2013)	The paper uses double DEA to evaluate the impact of accounting for quality on technical efficiency scores in 1,074 US hospitals. It compares various DEA quality indices and efficiency models, to determine how the treatment of quality influences findings regarding technical efficiency.	1,074 US hospitals.	Double DEA; FA;	Inputs: Beds, FTE Registered nurses, FTE Registered licensed practical nurses, FTE medical residents, Other staff, Percentage of AMI patients who received aspirin within 24 h of arrival, Percentage of AMI patients who received beta blocker within 24 h of arrival, Percentage of pneumonia patients with initial blood culture collected prior to antibiotics, Percentage of pneumonia patients who receive first antibiotics within 4 h of arrival. Outputs: Mortality rate within 30 days of discharge for acute myocardial infarction (AMI), Mortality rate within 180 days of discharge for Congestive Heart Failure, Mortality rate within 30 days of discharge for pneumonia, Surgical complication rate for cholecystectomy surgery (gall-bladder removal), Medicare IP days, Medicaid IP days, Other payer days, ER visits, Number of outpatients, Number of surgeries IP, Number of surgeries OP, Percentage of AMI patients prescribed aspirin at discharge, Percentage of AMI patients prescribed beta blocker at discharge, Heart Failure patients with Left Ventricular Systolic Dysfunction (LVSD) prescribed Angiotensin Converting Enzyme (ACE) inhibitor at discharge, Heart Failure patients with documented Left Ventricular Function (LVF) assessment or one scheduled at discharge.	The study highlights the importance of incorporating quality into efficiency studies, revealing significant variation in care quality across hospitals. It also highlights inefficiency in many hospitals and how adjusting for quality changes efficiency scores. The study concludes that efficiency scores are useful for identifying areas for improvement and recommends examining the implications of different quality adjustments and the distribution of quality-adjusted scores.

Nayar et al. (2013)	Measure hospital performance in terms of both technical efficiency and quality using DEA models. Explore the application of DEA as a benchmarking tool for quality in urban acute care hospitals in the US.	371 urban acute care hospitals.	DEA;	Inputs: Beds, FTE non-physician, Operational expenses, Average length of hospital stay. Outputs: Number of outpatients, Training FTE, Survival rate for AMI, Survival rate for CHF, Survival rate for pneumonia.	This study found that when both dimensions of performance, efficiency and quality, were considered, and a greater proportion of hospitals were found to be performing at the efficiency frontier. Less than 20% of the sample hospitals were optimally performing for both quality and efficiency. Tobit regression analysis of the DEA scores found that public, small, teaching hospitals had higher DEA efficiency and quality scores.
Castro et al. (2014)	This article presents a DEA model for Internal Medicine Services in most hospitals of the NHS, which evaluates the impact of services on global costs and restricts weight to avoid disproportionately heavy services. The study identifies complementary diagnostic and therapeutic methods and access to care as determinants of optimal performance.	48 hospitals in Portugal.	DEA;	Inputs: Doctors working time, Nurses working time, Expenses with nursing personnel, Expenses with exams, Expenses with drugs, pharmaceutical products and clinical consumables per severity-adjusted patient, Beds, Number of total hours of surgical interventions, Average length of hospital stay. Outputs: Number of patients discharged, Number of external consultations, Number of first outpatient appointments, Number of day hospital sessions, Number of day hospital users, Number of surgeries OP.	This study aimed to create a benchmarking platform for Portuguese hospitals to assess and improve their performance in internal medicine services. It found potential for efficiency improvement through cost savings in clinical supplies and medicines. Further research is needed to verify the approach with administrators and evaluate patient care efficacy. The project also aimed to operationalize evaluations with hospital managers.
Costa Jr et al. (2015)	This study aims to assess the operative time indicators in a public university hospital, through a descriptive cross-sectional study using data from operating room database.	12,114 procedures were performed from January 2011 to January 2012.	Cross-sectional descriptive study;	Inputs: Time of duration of anesthesia, Time of preparation of the OR (turnover), Time of duration of the operation, Time between the entrance of the patient into the OR and the beginning of anesthesia, Time between the beginning of anesthesia and the start of the surgery, Entrance of the first patient of the day into the OR, Lack of adequate material in the OR. Outputs: Number of cancelled surgeries, Number of late professionals, Number of surgeries produced, Time of the patient in the OR, Time between the end of the surgery and the end of the anesthesia, Time between the end of anesthesia and exit from the OR.	This study reveals the dynamic process of an operating room, highlighting the impact of non-surgical steps on productivity and efficiency. Although indicators can improve efficiency, delays and longer non-operative times negatively affect patient flow and resource utilization. Management and professional training could increase capacity.
Backman et al. (2016)	This paper aims to describe the successes and the challenges in the development of the Collaborative for Excellence in Healthcare Quality (CEHQ), an initiative to achieve higher quality of care in university hospitals in Canada.	52 teaching hospitals.	Delphi approach;	Inputs: Waiting time before surgery, Waiting time for diagnostic imaging, Waiting time in emergency department, Average length of hospital stay, Expected average length of stay. Outputs: Delay to hip fracture surgery, Percentage of alternate level of care cases/days, Readmission rate, Cost per weighted case, Number of pressure ulcer episodes per 100 episodes pertaining ulcers, Mortality rate, Surgical site infection (SSI), Surgical safety checklist, Vancomycin-resistant Enterococci (VRE), Clostridium difficile Infection (CDI), Methicillin-resistant Staphylococcus aureus (MRSA), Satisfaction with healthcare.	The Canadian Institute of Health Quality (CEHQ) is working to improve healthcare performance and quality data from Canadian hospitals. The pan-Canadian approach aims to maximize opportunities for peer-comparison and ensure valid representation from quality performance leaders. To improve reporting, strategies and best practices need to be explored, and data from all university hospitals must be generated on a comparable basis and accurately reflect quality of care. Collaboration with CIHI is needed to ensure accurate and timely data for benchmarking and overall health system performance improvement.

Chowdhury & Zelenyuk (2016)	Analyze production performance of hospital services in Ontario across geographic locations, size and teaching status, using a non-parametric DEA technique with bootstrapping and truncated regression.	Balanced panel data from 113 acute-care hospitals in Ontario for the years 2003 and 2006.	DEA; bootstrapping; Truncated regression;	<u>Inputs</u>: Administrative staff hours, Nurses working time, Beds, Expenses with medical-surgical supplies, Expenses with non-medical supplies, Beds occupancy rate, Expenses with equipment <u>Outputs</u>: Number of outpatients, Case-mix weighted inpatient days, Readmission rate, Logarithm of total revenue, Rate of unit producing personnel.	Several organizational characteristics, such as occupancy rate, rate of unit-producing personnel, outpatient–inpatient ratio, case-mix index, geographic locations, size and teaching status are significant determinants of efficiency, and small, rural, and non-teaching hospitals tend to be more effective in their operations. The assumption of separability, which is a drawback of the current strategy, may be overcome by new techniques investigated in future studies.
Portela et al. (2016)	Describe a benchmarking platform, called Hospital Benchmarking (HOBE), where hospitals' services are used as the unit of analysis.	80 indicators have been defined for 19 clinical services.	DEA;	<u>Inputs</u>: Expenses with external supplies and services, Expenses with medical doctors, Expenses with nursing personnel, Expenses with drugs, Expenses with equipment, Expenses with CDT (complementary means of diagnosis and therapy), Expenses with surgery wards, Average length of hospital stay, Average length of hospital stay in ICU (intensive care units). <u>Outputs</u>: Number of first outpatient appointments, Number of emergency appointments, Number of surgeries OP, Number of conventional surgeries, Number of urgent surgeries.	The authors argue that benchmarking hospitals is best by comparing services rather than entire institutions. HOBE models allow for aggregation of service performance and comparison between hospitals based on efficiencies. The authors suggest that HOBE could be extended to include a more thorough analysis of healthcare quality, which would improve its usefulness to the public and enable computation of performance measures accounting for both cost and service quality.
Flokou et al. (2017)	Apply the DEA method to measure the efficiency of Greek NHS hospitals between 2009– 2013. The study divided hospitals into four groups with common characteristics, allowing year-by-year comparisons.	107 Greek NHS hospitals.	DEA; Malmquist index;	<u>Inputs</u>: Beds, FTE physicians, Other staff, Average length of hospital stay. <u>Outputs</u>: Number of outpatients, Number of inpatients, Number of surgeries produced.	The study emphasizes the importance of identifying and eliminating inefficiencies in the healthcare sector to balance reduced budgets without compromising service quality. It provides an efficiency map for the Greek NHS from 2009 to 2013, emphasizing the need for specific criteria for measuring efficiency. Saved resources could be used for patient care, innovative technology, and staff training.
Khushalani & Ozcan (2017)	Examines efficiency of producing quality in hospitals between 2009 and 2013 using Dynamic DEA and the hospital characteristics that contribute to this efficiency.	1259 hospitals from US.	DNDEA (Dynamic network DEA);	<u>Inputs</u>: Beds, FTE non-physician, Total operating expenses per bed, Ratio of registered nurse FTEs to total nursing FTEs, Service mix of high technology services. <u>Outputs</u>: Number of unit discharges, case mix adjusted, Number of surgeries produced, Number of emergency appointments, Number of outpatients, % of hospital inpatients who would definitely recommend the hospital, % of hospital inpatients who gave the hospital an overall rating of nine and above out of ten, Mortality Rate within 30 days of discharge for Heart Attack, Readmission Rate within 30 days of discharge for Heart Attack, Mortality Rate within 30 days of discharge for Heart Failure, Readmission Rate within 30 days of discharge for Heart Failure, Net patient revenue per bed, Mortality rate within 30 days of discharge for pneumonia, Readmission rate within 30 days after discharge for pneumonia.	The efficiency of both the quality and medical/surgical patient care sub-units improved significantly between 2009 and 2013 in US hospitals. Non-teaching and rural hospitals showed more improvement in efficiency than teaching and urban hospitals, which goes against Resource Dependence Theory. The study also found that improvements in efficiency in both sub-units are positively correlated, which means that there are no trade-offs between improving quality and efficiency. The study has important implications for policymakers, practitioners, and DEA researchers, but it is limited by the exclusion of some quality measures and the potential impact of the ACA provisions.

Ferreira et al. (2018a)	Analysis of the scale efficiency, optimal scale for hospital clinical staff, and the exogenous dimensions that can be associated with them.	108 Portuguese hospitals-year.	DEA; PCA;	Inputs: Doctors per 1000 inhabitants, FTE Registered nurses, Beds, Operational expenses, Waiting time before surgery, Average length of hospital stay. Outputs: Number of inpatients, Number of medical appointments, Operating theatre, Number of emergency appointments, Number of surgeries produced, Number of emergency episodes, Number of unit discharges, case mix adjusted, Number of first medical appointments within the legally defined time, Number of outpatient scheduled surgeries, Number of urgent surgeries, Number of outpatient surgeries per 100 potential outpatient procedure, Caesarean section rate per delivery, Readmissions within 30 days after discharge per 100 inpatients, Obstetric trauma rate, Postoperative septicaemia rate, Postoperative pulmonary embolism/ deep vein thrombosis rate, Mortality rate for less severe cases, Number of first nonurgent medical appointments within the legally defined time per 100 first nonurgent medical appointments, Number of hip surgeries in the first 48 h after fracture per 100 hip surgeries.	This study highlights the importance of optimizing the allocation of clinical staff in Portuguese hospitals to improve scale efficiency without compromising social values. Hospitals located in the Portuguese coastline are mostly operating above their optimal scale size, which leads to a misdistribution of health workforce. Health policies should aim to reduce or increase the workforce in hospitals operating above or below their optimal size, respectively, in order to improve scale efficiency.
Kohl et al. (2019)	By examining the research purposes of 262 publications of DEA applications in healthcare, this paper aims to be of assistance to researchers planning to apply DEA in a hospital. It analyzes the methodological settings of the studies and describes the applied models, as well as the chosen inputs and outputs and the relevant downstream techniques.	262 publications of DEA applications in healthcare.	DEA; Malmquist index;	Inputs: Beds, FTE physicians, FTE Registered nurses, FTE Registered licensed practical nurses, FTE non-physician, FTE Other staff, Material supplies, Gross equipment, Total cost, , Other costs, Doctors working time, Nurses working time, Expenses with medical-surgical supplies, Waiting times. Outputs: Number of outpatients, Number of inpatients, Number of surgeries produced, Number of services offered, Average doctor visits/hour, Mortality rate, Readmission rate, Number of medication errors, Intra-hospitals infection rate, Service and performance	The research purpose is divided into four clusters: 36 investigate the effects of reforms, 48 introduce new methodologies, 100 answer management questions, and 99 focus on efficiency estimation. The study identifies a trend towards using quality parameters in model specification, but caution is needed when selecting models. The DEA model is chosen to estimate efficiency scores in 80% of publications, but there is no proof of its reliability.
Gomes et al. (2020)	Evaluate and compare the quality of care in operating rooms in Portuguese hospitals regarding their structure, process, and outcome.	71 Portuguese hospitals.	Questionnaire;	Gender; Region; Activity of the operating room; Environment and equipment; Resources for quality and safety; Circuits in the operating room; Facilities and operating requirements; Training and praxis in the operating room; Continuity in nursing care; Specificities of the professional groups; Communication flow; Support strategies to processes of care work; Control for safety; Teamwork; Evaluation of the quality of care; Ownership status; Evaluation of processes; Control systems.	The study assessed operating room standards in Portuguese hospitals, finding that charity, commercial, and governmental hospitals had the best structure and process values, however public institutions had lower results. The report recommends more research involving private institutions and using indicators to track healthcare for quality and safety.

Ferreira & Marques (2021)	Use of robust benchmarking methodologies alongside recent data about Portuguese hospitals (2012-2017) to demonstrate that PPP hospitals can deliver health care services with social performance levels at least as good as public hospitals.	2,660 observations of hospitals: 2,380 of public hospitals and 280 of hospital PPP arrangements.	BoD;	Inputs: Beds, Waiting time before surgery. Outputs: Number of outpatient surgeries per 100 potential outpatient procedure, Readmission rate within 30 days after discharge, Rate of inpatients staying for more than 30 days, Decubitus ulcer cases per 1000 inpatients, Catheter related bloodstream infections per 1000 inpatients, Number of postoperative pulmonary embolism/deep vein thrombosis cases per 100000 inpatients, Postoperative septicaemia cases per 100 inpatients, First non-urgent medical appointments within the maximum guaranteed time rate, Number of hip surgeries in the first 48 h after fracture per 100 hip surgeries.	The study finds no reason for believing that public hospitals outperform PPPs in terms of social performance. Both groups have high levels of social inefficiency, notably in areas like timeliness and care appropriateness. Nevertheless, other dimensions like structural quality and outcomes are equally significant and are not considered in this study, only considering aspects relating to the process of care. The paper advises utilizing a bias-/heterogeneity-corrected input-oriented PPP model with financial inputs and raw outputs to benchmark and compare the two sets of hospitals.
Matos et al. (2021)	Use of the Benefit of Doubt methodology alongside recent data about Portuguese public hospitals (2013–2017) to understand the factors that contribute to low performance and high indebtedness levels. Analyze Portuguese public hospitals' overall performance as measure by a global CI based on four other CIs (access, efficiency and productivity, financial, and quality).	5 single hospitals and 18 HCs in Portugal.	BoD; PCA;	Inputs: Average length of hospital stay, Beds occupancy rate, Standard patient per FTE nurse, Standard patient per FTE doctor, Total operating expenses per bed, Expenses with staff compensation, Expenses with external supplies and services, FTE physicians, FTE Registered nurses, Drug expenses per standard patient, Waiting time before surgery. Outputs: Number of hip fracture surgery in the first 48 h, Number of first medical appointments within the legally defined time, Number of surgeries produced within time, Caesarean section rate (without justification), Number of outpatient surgeries on potential outpatient procedures, Rate of inpatients staying for more than 30 days, Readmission rate within 30 days after discharge, Postoperative pulmonary embolism/ deep vein thrombosis rate, Postoperative septicaemia rate, Number of cases of trauma on vaginal delivery (instrumented and non-instrumented) with lacerations of third and fourth degree.	The emphasizes the need for improvement in financial aspects to boost efficiency and productivity. Healthcare providers must balance access, efficiency, productivity, financial, and quality groups to achieve higher performance levels. The study uses a tool for performance assessment, ranking hospitals and motivating improvements; however, the results are not definitive, and further research is needed to validate the findings, including incorporating ROE and ROI indicators and exogenous factors. The compensatory nature of classic BoD was observed, but weight restrictions reduced this effect. The authors recommend the use of the order-m model and suggest penalties for better discrimination of results.

Pereira et al. (2021a)	The authors propose a novel approach based on the directional 'Benefit-of-the-Doubt' model, to construct a Composite Indicator that assesses the Portuguese public hospitals' performance under two perspectives of hospital activity: users and providers, using weight restrictions and an artificial target.	29 public hospitals in Portugal.	MCDA; DEA;	Inputs: Standard patient per FTE physician, Standard patient per FTE nurse, Inpatient bed occupancy rate, Expenses with staff per severity-adjusted patient, Expenses with drugs, pharmaceutical products and clinical consumables per severity-adjusted patient, Expenses with materials consumed per standard patient, Expenses with supplies and external services per severity-adjusted patient, Waiting time before surgery. Outputs: Number of first nonurgent medical appointments within the legally defined time per 100 first nonurgent medical appointments, Number of registers on surgery waiting list beyond the legally defined time, Number of outpatient surgeries per 100 potential outpatient procedure, Number of hip surgeries in the first 48 h after fracture of users above 65y per 100 users over 65y with surgically repaired hip fractures, Number of readmissions in 30 days after discharge per 100 inpatients, Number of long-stay inpatients per 100 admissions (30 days), Number of caesarean sections per 100 deliveries, Number of pressure ulcer episodes per 100 episodes pertaining ulcers, Bloodstream infection rate related to CVC per 100 inpatients, Number of postoperative pulmonary embolism/deep vein thrombosis cases per 100000 inpatients, Number of post-operative sepsis episodes per 100,000 episodes pertaining possible post-operative sepsis, Number of cases of assisted vaginal deliveries with 3rd and 4th degree lacerations per 100 deliveries, Number of cases of non-assisted vaginal deliveries with 3rd and 4th degree lacerations per 100 deliveries.	The findings indicate that only a small number of hospitals in the Portuguese NHS are effective, and the dearth of effective public hospitals continues to be a serious problem. The policy repercussions deduced from these findings emphasize the necessity of immediate interventions in the NHS reform. Further study options include evaluating the performance of both the primary and secondary healthcare providers in the NHS over time and incorporating preference data from numerous DMs into DEA models using alternative MCDA methodologies.
Pereira et al. (2021b)	Measuring the efficiency of the NHS's secondary healthcare providers using a value modelled slacks-based network DEA with simulation applied to a matrix-type structure.	27 healthcare institutions operating as EPE in Portugal.	DEA;	Inputs: Expenses with staff compensation, Total cost, Clinical material costs, Expenses with external supplies and services. Outputs: Number of patients discharged, Number of scheduled surgeries, Number of critical surgeries, Number of minor OP surgeries, Number of medical appointments, Number of emergency appointments.	A study using a value-modelled network DEA and simulation to measure the efficiency of secondary healthcare providers in Portugal's Social Networks (NHS) revealed that 2/3 of these providers were inefficient, contrasting with recent "black-box DEA" studies that only 40% of public hospitals were deemed inefficient. The majority of efficient hospitals were located in coastal areas, with the most frequent efficient services being inpatient and operating room. Outpatient surgeries, medical appointments, and emergency services were often inefficient, even after 5000 iterations of the Monte Carlo simulation method.

Rocha et al. (2021)	This work aims to assess the quality of Portuguese public hospitals using a Multiple Criteria Decision Aiding (MCDA) method – Electre Tri-nC.	25 Hospitals in Portugal.	MCDA - Electre Tri-nC;	Inputs: First medical appointments timeliness, Beds occupancy rate, Doctors per 1000 inhabitants, Nurses per 1000 inhabitants, Expenses with staff per severity-adjusted patient, Expenses with drugs, pharmaceutical products and clinical consumables per severity-adjusted patient, Expenses with supplies and external services per severity-adjusted patient, Expenses with overtime per total expenses with staff, Expenses with outsourcing per total expenses with staff, Waiting time before surgery. Outputs: Number of outpatient surgeries per 100 potential outpatient procedure, Number of readmissions in 30 days after discharge per 100 inpatients, Number of long-stay inpatients per 100 admissions, Number of hip surgeries performed in the first 48 hours per 100 hip surgeries, Number of bedsores per 100 inpatients, Bloodstream infection rate related to CVC per 100 inpatients, Number of postoperative pulmonary embolism/deep venous thrombosis cases per 100 surgical procedures, Postoperative septicaemia cases per 100 inpatients, Number of cases of trauma on vaginal delivery (third and fourth degree lacerations), without instrumentation, per 100 assisted deliveries, Number of cases of trauma on vaginal delivery (third and fourth degree lacerations), with instrumentation, per 100 assisted deliveries, Number of caesarean sections per 100 deliveries, Number of caesarean sections in UCFTPs per 100 sections in UCFTPs, Number of first caesarean sections in UCFTPs per 100 deliveries in UCFTPs without caesarean section before, Enrolled patients for surgery.	The results showed that no hospitals were assigned to the best category, and several were assigned to the worst category. In 2018, 60% of hospitals were assigned to a category equal or less than the worst performance category. Centro Hospitalar de Póvoa do Varzim/Vila do Conde was considered the best hospital in both years and views, suggesting the possibility of using this entity as a benchmark. Only 8% of assignments could change, indicating the results are robust. However, the study's results depend on criteria preference, which is developed jointly between the analyst and the DM. Access to other DMs could provide information about the model's adaptability to other opinions and scenarios.
Lin et al. (2022)	Estimate the efficiency of 19 tertiary hospitals in Taiwan using a two-stage analysis of DEA and TOBIT regression.	19 tertiary hospitals in Taiwan.	DEA; Tobit regression;	Inputs: Beds, FTE physicians, Net Value of fixed assets, Gross equipment, Self-pay income, Average length of hospital stay. Outputs: Surplus or deficit of appropriation, Total relative value units [RVUs] for outpatient services, Total relative value units [RVUs] for inpatient services, Modified EBITDA, Rate of emergency transfer inpatient stay over 48h.	The study reveals that private hospitals have better BCC pure technical efficiency (PTE) than public hospitals, while public hospitals have superior CCR (TE). The highest correlation is found in hospital beds and days. The CCR and BCC models show similar efficiency over four years, but more hospitals need improvement in surplus or deficit of appropriation, modified EBITDA, and self-pay income.
Nunes & Ferreira (2022a)	The concept of "health in all policies" and its impact on sustainability, innovation, and social inequalities in health. Construction of a composite indicator useful to evaluate the quality of life of citizens.	277 Portuguese municipaliti es (located in Portugal mainland).	DEA; BoD;	Inputs: Crude birth rate, Doctors per 1000 inhabitants, Pharmacists per 1000 inhabitants. Outputs: Mortality rate.	Observed regional asymmetries with more significant evidence regarding health and living conditions, but uniformity in environmental issues and education and culture. The considerable heterogeneity in health (and the overall quality of life) seems to be justified, in part, by geographic variables.

Nunes & Ferreira (2022b)	Development of a network DEA model relating to efficiency and effectiveness, to evaluate the performance of hospital units before and during the pandemic period.	37 Portuguese public hospitals (secondary care) containing data for 18 variables.	Network DEA;	Inputs: Beds, Days before surgery, Doctors working time, Nurses working time, Average length of hospital stay. Outputs: Number of standard patients, Number of first medical appointments within the legally defined time, Number of registers on the surgery waiting list within the legally defined time, Number of hip surgeries in the first 48 h after fracture, Number of inpatients without adverse events, Number of delayed first medical appointments, Number of registers on surgery waiting list beyond the legally defined time, Number of readmitted cases after 30 days of inpatient discharge, Number of hip surgeries beyond 48 h after fracture, Number of avoidable bedsore cases, Number of avoidable bloodstream infections, Number of avoidable postoperative pulmonary embolisms, Number of septicaemia cases.	The study examines the efficiency and effectiveness of hospital healthcare providers in Portugal before and during the COVID-19 pandemic. Efficiency is crucial for service sustainability, patient safety, and compliance with medical guidelines. In 2020, efficiency decreased due to decreased hospital production, but there was no significant decrease in input. During the pandemic, efficiency maintained a growing trend, possibly due to the exodus of highly qualified health professionals seeking better salaries and working conditions. Access barriers, such as postponements and cancellations of appointments, were evident. Vaccination reduced severe COVID-19 cases in 2021, but other undesirable cases could have increased, affecting hospital care effectiveness. Hospital readmissions and timely hip surgeries offset poor results on quality variables.

Table 9- Results of the Shapiro-Wilk normality test for performance indicators, in 2019 (Source: Author).

	W	p-value
q1	0,488	0,000
q2	0,842	0,000
q3	0,905	0,004
q4	0,976	0,612
q5	0,964	0,285
q6	0,982	0,806
q7	0,981	0,778
q8	0,891	0,002
q9	0,939	0,044
q10	0,877	0,001

Table 10- Results of the Shapiro-Wilk normality test for performance indicators, in 2020 (Source: Author).

	W	p-value
q1	0,983	0,819
q2	0,822	0,001
q3	0,973	0,485
q4	0,974	0,524
q5	0,96	0,206
q6	0,988	0,946
q7	0,978	0,654
q8	0,924	0,015
q9	0,989	0,969
q10	0,847	0,001

Table 11- Results of the Shapiro-Wilk normality test for performance indicators, in 2021 (Source: Author).

	W	p-value
q1	0,975	0,567
q2	0,680	0,001
q3	0,198	0,001
q4	0,982	0,809
q5	0,963	0,260
q6	0,981	0,779
q7	0,964	0,279
q8	0,914	0,007
q9	0,990	0,976
q10	0,905	0,004

Table 12- Pearson correlation values between performance indicators (Source: Author).

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10
q1	1,000	-0,062	-0,090	-0,258	0,149	0,011	0,061	-0,053	-0,047	-0,071
q2	-0,061	1,000	-0,044	0,214	0,190	0,296	-0,529	0,190	-0,002	-0,293
q3	-0,089	-0,044	1,000	-0,201	-0,134	0,220	0,318	0,066	0,270	-0,017
q4	-0,258	0,214	-0,201	1,000	0,395	-0,198	-0,159	0,113	-0,275	0,081
q5	0,149	0,190	-0,134	0,395	1,000	-0,440	-0,336	0,037	-0,068	-0,046
q6	0,012	0,297	0,220	-0,198	-0,440	1,000	-0,052	0,117	0,124	-0,121
q7	0,061	-0,529	0,318	-0,159	-0,336	-0,052	1,000	-0,221	0,122	0,304
q8	-0,053	0,190	0,066	0,113	0,037	0,117	-0,221	1,000	0,005	-0,076
q9	-0,047	-0,002	0,270	-0,275	-0,068	0,124	0,122	0,005	1,000	-0,141
q10	-0,071	-0,293	-0,017	0,081	-0,046	-0,121	0,304	-0,076	-0,141	1,000

Table 13- Average values of performance indicators for top performing and under-performing units (Source: Author).

		q1	**q2**	**q3**	**q4**	**q5**	**q6**	**q7**	**q8**	**q9**	**q10**
2019	Top performing (CI >= 0,9)	90,31%	20,53%	69,61%	69,95%	80,14%	63,73%	2,99%	5,68%	8,70%	0,03%
	Under-performing (CI < 0,9)	88,99%	14,01%	72,80%	68,18%	77,20%	65,05%	4,25%	6,90%	10,44%	0,05%
2020	Top performing (CI >= 0,9)	75,20%	23,46%	82,35%	72,97%	87,22%	67,35%	3,61%	5,40%	8,94%	0,02%
	Under-performing (CI < 0,9)	76,35%	12,39%	79,08%	57,15%	75,24%	62,31%	4,53%	6,28%	9,32%	0,09%
2021	Top performing (CI >= 0,9)	83,41%	24,67%	77,78%	89,72%	91,30%	70,80%	3,40%	5,15%	8,22%	0,05%
	Under-performing (CI < 0,9)	76,57%	12,77%	73,43%	71,25%	76,81%	64,67%	4,24%	6,67%	8,54%	0,11%

Table 14- Percentage of time that entities were considered as benchmarks (Source: Author).

Hospital	% times (CI = 1)
CHBM	0,0%
CHL	2,8%
CHLO	0,0%
CHS	0,0%
CHBV	5,6%
CHMA	38,9%
CHO	0,0%
CHUC	2,8%
CHEDV	38,9%
CHMT	0,0%
CHPVVC	52,8%
CHTS	8,3%
CHTV	0,0%
CHTMAD	0,0%
CHUCB	0,0%
CHULC	0,0%
CHUSJ	0,0%
CHUA	0,0%
CHUP	33,3%
CHULN	0,0%
CHVNGE	5,6%
HSOG	0,0%
HB	33,3%
HDFF	44,4%
HDS	13,9%
HESE	2,8%
HPDFF	0,0%
HGO	2,8%
HSMM	44,4%
ULSG	5,6%
ULSCB	5,6%
ULSM	13,9%
ULSAM	33,3%
ULSBA	16,7%
ULSLA	5,6%
ULSN	22,2%
ULSNA	0,0%

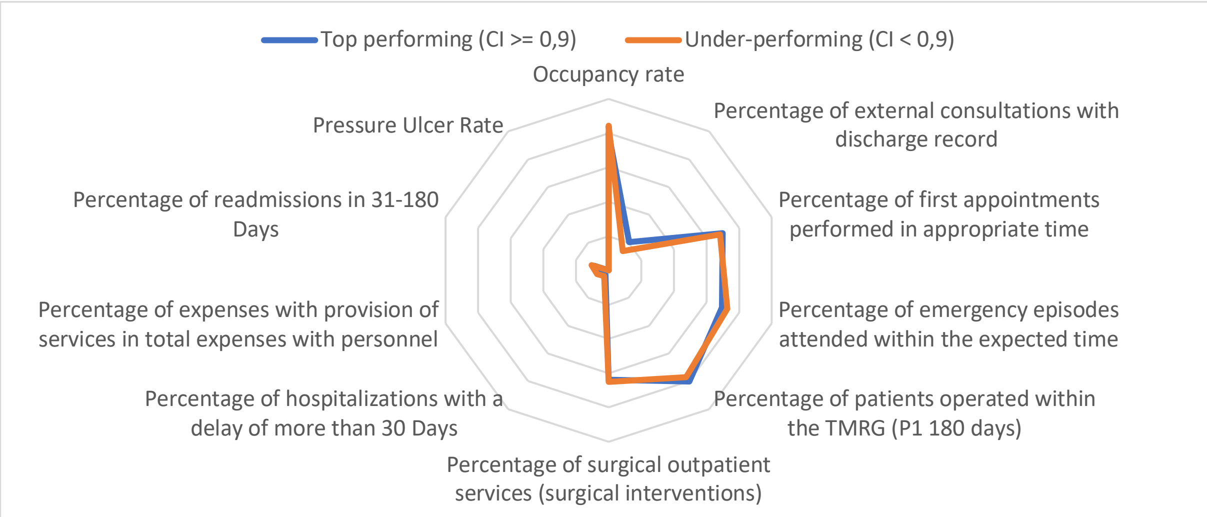

Figure 11- Comparison between average values of top performing and under-performing units in 2019 (Source: Author).

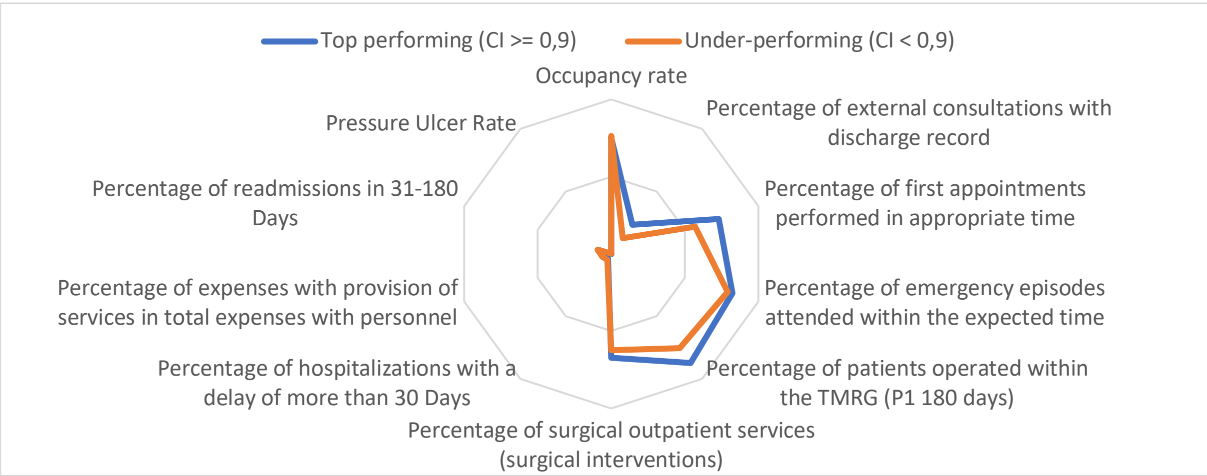

Figure 12- Comparison between average values of top performing and under-performing units in 2020 (Source: Author).

Table 15- CI values obtained according to the restriction for the minimum weight limits (Source: Author).

Minimum weight	**0**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**	**0,1**
CHBM	1,00	0,86	0,79	0,76	0,70	0,64	0,59
CHL	1,00	0,24	0,20	0,18	0,16	0,14	0,13
CHLO	0,98	0,67	0,57	0,50	0,44	0,40	0,36
CHS	1,00	0,64	0,53	0,48	0,43	0,39	0,36
CHBV	1,00	0,99	0,92	0,90	0,89	0,83	0,77
CHMA	1,00	0,99	0,96	0,93	0,89	0,82	0,76
CHO	1,00	0,69	0,62	0,58	0,54	0,51	0,47
CHUC	1,00	0,85	0,80	0,74	0,66	0,60	0,55
CHEDV	1,00	0,93	0,90	0,88	0,85	0,80	0,74
CHMT	0,99	0,68	0,58	0,54	0,51	0,48	0,45
CHPVVC	1,00	0,00	0,00	0,00	0,00	0,00	0,00
CHTS	1,00	0,91	0,57	0,53	0,49	0,45	0,42
CHTV	1,00	0,00	0,00	0,00	0,00	0,00	0,00
CHTMAD	1,00	0,00	0,00	0,00	0,00	0,00	0,00
CHUCB	0,99	0,74	0,68	0,64	0,60	0,54	0,50
CHULC	1,00	0,44	0,36	0,32	0,28	0,25	0,23
CHUSJ	0,97	0,55	0,47	0,41	0,36	0,33	0,30
CHUA	0,98	0,00	0,00	0,00	0,00	0,00	0,00
CHUP	1,00	0,95	0,95	0,92	0,87	0,80	0,73
CHULN	1,00	0,00	0,00	0,00	0,00	0,00	0,00
CHVNGE	1,00	0,71	0,66	0,60	0,54	0,49	0,45
HSOG	1,00	0,00	0,00	0,00	0,00	0,00	0,00
HB	1,00	1,00	1,00	0,97	0,88	0,81	0,74
HDFF	1,00	0,97	0,92	0,88	0,84	0,80	0,75
HDS	1,00	0,92	0,86	0,82	0,78	0,75	0,70
HESE	1,00	0,90	0,83	0,77	0,69	0,63	0,58
HPDFF	1,00	0,00	0,00	0,00	0,00	0,00	0,00
HGO	1,00	0,88	0,81	0,77	0,70	0,64	0,59
HSMM	1,00	1,00	1,00	1,00	0,94	0,86	0,80
ULSG	1,00	0,65	0,59	0,53	0,48	0,43	0,40
ULSCB	1,00	0,94	0,91	0,88	0,84	0,80	0,74
ULSM	1,00	1,00	1,00	0,97	0,93	0,85	0,78
ULSAM	1,00	1,00	1,00	1,00	0,99	0,96	0,89
ULSBA	1,00	0,97	0,92	0,90	0,88	0,86	0,81
ULSLA	1,00	0,00	0,00	0,00	0,00	0,00	0,00
ULSN	0,94	0,87	0,88	0,86	0,84	0,81	0,78
ULSNA	1,00	0,60	0,49	0,44	0,41	0,37	0,34